Structure and Function of Human Autophagy Protein ATG16L1

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig

zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)

genehmigte

Dissertation

von Milica Bajagić
aus Podgorica, Montenegro

1. Referent:	Professor Dr. Michael Steinert
2. Referent:	Professor Dr. Wulf Blankenfeldt
eingereicht am:	04.11.2015
mündliche Prüfung (Disputation) am:	05.07.2016

Druckjahr 2017

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

1. Tagungsbeiträge

Bajagic, M., Leupold, S., Bleckmann, M., Walker A., Schmitz, I., Van den Heuvel, J. and Scrima, A. Autophagy in immunity and infection. 5th Annual PhD Retreat of the Helmholtz International Graduate School, Hahnenklee, Goslar, Germany (May 2014)

Bajagic, M., Van den Heuvel, J. and Scrima, A. Structure and function of human autophagy protein ATG16L1. 7th International PhD symposium of the Helmholtz International Graduate School, Helmholtz Center for Infection Research, Braunschweig, Germany (December 2014)

Bajagic, M., Van den Heuvel, J. and Scrima, A. Structure and function of human autophagy protein ATG16L1. FEI Cryo-TEM workshop, Leiden, Netherlands (March 2015)

2. Posterbeiträge

Bajagic, M., Van den Heuvel, J. and Scrima, A. Towards understanding the mechanism of autophagy-mediated intracellular pathogen clearance. 3rd Annual PhD Retreat of the Helmholtz International Graduate School, Gustav Stresemann Institute in Niedersachsen, Bad Bevensen, Germany (October 2012)

Bajagic, M., Van den Heuvel, J. and Scrima, A. Towards understanding the mechanism of autophagy-mediated intracellular pathogen clearance. 4rd Annual PhD Retreat of the Helmholtz International Graduate School, Hahnenklee, Goslar, Germany (May 2013)

Bajagic, M., Van den Heuvel, J. and Scrima, A. Structure and function of human autophagy protein ATG16L1. 6th International PhD symposium of the Helmholtz International Graduate School, Helmholtz Center for Infection Research, Braunschweig, Germany (December 2013)

Bajagic, M., Van den Heuvel, J. and Scrima, A. Structure and function of human autophagy protein ATG16L1. Jürgen Wehland Symposium “North Regio Day on Infection” (NoRDI) V, Helmholtz Center for Infection Research, Braunschweig, Germany (October 2014)

Table of Contents

List of Tables

List of Figures

Abbreviations

Å	Ångström; 1Å = 10^{-10} m
AP	Alkaline phosphatase
APS	Ammonium Persulfate
ATP	Adenosine triphosphate
ATG	Autophagy-related protein
AU	Absorption unit
BME	2-Mercaptoethanol
CD	Circular dichroism
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CV	Column volume
Da	Dalton; 1Da = 1g/mol
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleosid-triphosphate
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
Frw	Forward
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase

IPTG	Isopropyl-β-D-thiogalactopyranosid
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation – Time Of Flight
MDP	Muramyl dipeptide
MOPS	3-(N-morpholino)propanesulfonic acid
MR	Molecular replacement
MS	Mass spectrometry
MWCO	Molecular weight cutoff
Ni-NTA	Nickel-nitrilotriacetic acid
OD_{600}	Optical density at wavelength of 600nm
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein database
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PMSF	Phenylmethanesulfonylfluoride
PVDF	Polyvinyl difluoride
rev	Reverse
r.m.s.d.	Root main square deviation
rpm	Revolutions per minute
S75	Superdex 75 (gel filtration column)
S200	Superdex 200 (gel filtration column)
SDS	Sodium dodecyl sulfate

SDS PAGE	SDS polyacrylamide gel electrophoresis
SOC	Super Optimal broth with Catabolite repression
TEMED	Tetramethylethylenediamine
TEV	Catalytic domain of the nuclear inclusion protein (NIa) of the tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
(v/v)	Volume per volume
(w/v)	Weight per volume

Synopsis

Autophagy is a conserved eukaryotic mechanism for the sequestration and degradation of cellular components. Upon nutrient depletion or stress, a double-membrane vesicle referred to as autophagosome engulfs a section of cytoplasm and delivers it to lysosome for degradation. Apart from its classical role, autophagy is involved in human immunity, especially in capturing and clearance of intracellular pathogens. One of the key players in autophagy, ATG16L1, essential for autophagosome formation and elongation, recently came into focus of interest due to mutations in the gene locus that result in increased susceptibility to the inflammatory bowel disease Morbus Crohn. ATG16L1, in complex with autophagy-related proteins ATG5 and ATG12, localizes on the autophagosome and has also been shown to be recruited by the cytosolic pattern recognition receptors NOD1 and NOD2 to the entry site of cell invading bacteria, thus linking autophagy with the intracellular bacterial sensing.

Autophagy has been extensively studied in yeast, and more than 36 proteins have been identified as important for this process. Most of yeast autophagy proteins have mammalian homologues, including ATG16. Yeast ATG16 has an N-terminal ATG5 binding domain (ATG5 BD), and a coiled coil domain (CCD) responsible for dimerization. Human ATG16L1 also has these two domains, which are essential and sufficient for functional canonical autophagy. In addition to ATG5 BD and the CCD, human ATG16L1 has a C-terminal domain which is absent in its yeast homologue. Precisely this domain was shown to interact with NOD1 and NOD2, in addition to TLR2, TMEM59 and DDB1. In the course of this work, various biophysical methods were used to study these interactions using purified proteins. Under the experimental conditions tested, these interactions could not be detected. Further work is needed to identify possible factors missing in the experimental set up and detect mediating factors essential for macromolecular interactions with the C-terminal domain of the human ATG16L1.

C-terminal domain of human ATG16L1 was used for crystallographic studies and a high resolution structure of this domain was solved. Structure of the C-terminal domain of human ATG16L1 revealed that it folds into a typical WD-40 propeller with seven blades, each composed of four anti-parallel β-strands. The WD-40 proteins are a large family found in all eukaryotes and usually serve as interaction platform for peptides, nucleic acids and other proteins. Detailed analysis of the structure identified a region on the WD-40 domain of ATG16L1 with high flexibility and a possible functional relevance. Additionally, two symmetry related molecules of the WD-40 domain of ATG16L1 bury a large surface area, which might be physiologically important in mediating dimerization of the WD-40 domains. N-termini of these two molecules point in the same direction, indicative of an orientation in agreement with a parallel dimer formed via a CCD of the ATG16L1. Based on these observations, the WD-40 domain of ATG16L1 might form a WD-40/WD-40 interaction surface for other macromolecules and peptides, making this domain an interesting candidate for an intersection between autophagy and other cellular processes.

Synopsis

Autophagie ist ein konservierter eukaryotischer Mechanismus für die Sequestrierung und den Abbau von intrazellulären Komponenten. Bei Nährstoffverarmung oder Stress umschließt ein Doppel-Membran-Vesikel, welches als Autophagosom bezeichnet wird, einen Teil des Cytoplasmas und transportiert den Inhalt weiter zum Abbau durch Lysosomen. Abgesehen von ihrer klassischen Rolle ist Autophagie an zahlreichen Prozessen im Kontext der menschlichen Immunität beteiligt, insbesondere an der Erkennung und Beseitigung von intrazellulären Pathogenen. Einer der wichtigsten Akteure in Autophagie, ATG16L1, das für die Ausbildung und Reifung des Autophagosoms bedeutend ist, rückte vor kurzem in den Fokus des Interesses aufgrund von Mutationen im Gen-Locus, welche zu einer erhöhten Anfälligkeit für die entzündliche Darm-erkrankung Morbus Crohn führen. ATG16L1, im Komplex mit den Autophagie verwandten Proteinen ATG5 und ATG12, ist auf dem Autophagosom lokalisiert und wird durch die cytosolischen Muster erkennenden Rezeptoren NOD1 und NOD2 zur Eintrittstelle von invasiven Bakterien rekrutiert, was eine direkte Verbindung zwischen Autophagie und der Erkennung intrazellulärer Bakterien aufzeigt.

Autophagie wurde ausgiebig in Hefe untersucht und mehr als 36 für diesen zellulären Prozess essentielle Proteine wurden identifiziert. Die Mehrzahl dieser Proteine hat Säugerhomologe, einschließlich ATG16. Hefe-ATG16 hat eine N-terminale ATG5 Bindungsdomäne (ATG5 BD) und eine Coiled-Coil-Domäne (CCD), welche für die Dimerisierung verantwortlich ist. Menschliches ATG16L1 besitzt diese Domänen ebenfalls, welche für eine funktionierende kanonische Autophagie notwendig und ausreichend sind. Neben ATG5 BD und der CCD besitzt menschliches ATG16L1 eine C-terminale Domäne, die im Hefe Homolog des ATG16L1 fehlt. Gerade bei dieser Domäne wurde eine Interaktion mit NOD1 und NOD2, neben TLR2, TMEM59 und DDB1 nachgewiesen. Im Rahmen dieser Arbeit wurden verschiedene biophysikalische Methoden verwendet, um diese Interaktionen mit gereinigten Proteinen zu untersuchen. Unter den hier verwendeten Versuchsbedingungen konnten diese Wechselwirkungen jedoch nicht bestätigt werden. Weitere Arbeiten sind erforderlich, um eventuelle zusätzliche Faktoren zu ermitteln, welche als verbrückendes Bindeglied für die makromolekularen Wechselwirkungen mit der C-terminalen Domäne des menschlichen ATG16L1 essentiell sind.

Zusätzlich wurde die C-terminale Domäne des menschlichen ATG16L1 für kristallographische Untersuchungen verwendet und die Struktur dieser Domäne aufgeklärt. Die hochaufgelöste Struktur der C-terminalen Domäne von menschlichem ATG16L1 zeigt, dass die Faltung der Domäne der eines WD-40 Propeller mit sieben Propeller-Blättern entspricht, wobei sich jedes Blatt aus vier antiparallelen β-Strängen zusammensetzt. Die WD-40-Proteine sind eine große Familie, die in allen Eukaryoten präsent ist und in der Regel als Interaktionsplattform für Peptide, Nukleinsäuren und andere Proteine dient.

Durch eine detaillierte Analyse der Struktur wurde eine Region auf der WD-40-Domäne des ATG16L1 mit hoher Flexibilität und einer möglichen funktionellen Bedeutung identifiziert. Zusätzlich wurde eine durch symmetrie-verwandte Moleküle der WD-40-Domäne generierte Interaktionsfläche identifiziert, die eine Dimerisierung der WD-40 vermitteln und von physiologischer Relevanz sein könnte. Die N-Termini dieser zwei Moleküle weisen in die gleiche Richtung, was auf eine Ausrichtung in Übereinstimmung mit einem parallelen Dimer hindeutet, das über die CCD des ATG16L1 gebildet wird. Auf Basis dieser Beobachtungen könnte die WD-40 Domäne von ATG16L1 eine WD-40/WD-40 Interaktionsfläche für andere Makromoleküle und Peptide bilden, so dass diese Domäne einen interessanten Kandidaten für die Funktion als Bindeglied zwischen Autophagie und anderen zellulären Prozessen darstellt.

1 Introduction

1.1 Cell maintenance and turn over mechanisms

All living organisms on Earth depend on three basic and essential macromolecules for their biological functions: DNA, RNA and proteins. Proteins are large biopolymers, generated as a string of a combination of 20 essential amino acids which can fold into different three dimensional moieties, generally called domains. Proteins can be further diversified by posttranslational modifications or binding to cofactors, other proteins, or nucleic acids. The amount of a protein in the cell at any given time is determined by two factors: rate of its synthesis and rate of its degradation. The degradation rate of proteins (and other cellular components) is a crucial step in cellular regulation. Depending on their function, some proteins must be synthesized and degraded quickly to adapt to ever changing cellular environment and extracellular stimuli. Other proteins are degraded upon reception of certain signals, to prevent overstimulation. Due to a vast amount of proteins being manufactured in the cell at any given time, some of them have errors leading to protein misfolding or their aggregation. Such proteins must be rapidly degraded. Eukaryotic cells have two major pathways responsible for degradation of faulty proteins: proteasome pathway and lysosome pathway (Figure 1).

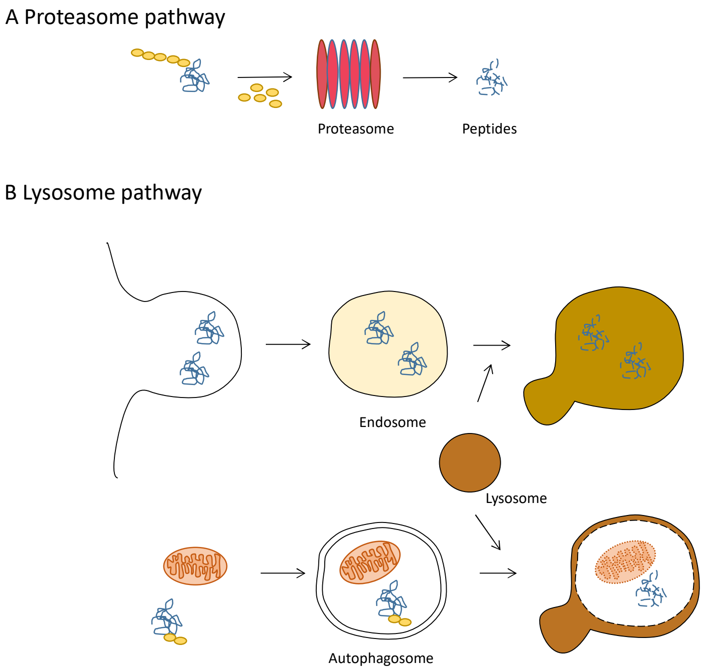


Figure 1: Two main routes of intracellular degradation in eukaryotic systems. A) Proteasome pathway. Proteins are targeted to the proteasome via ubiquitin tag (yellow circles). B) Lysosome pathway. Extracellular and membrane proteins are taken up by endocytosis, while large intracellular aggregates and damaged organelles are taken up by autophagy, and delivered to the lysosome for degradation.

Proteasome is a large cylindrical complex of proteases responsible for recycling of misfolded proteins. Protein folding is often facilitated by chaperons, which ensure proper environment for the polypeptide chain to assume its three dimensional structure. If the protein is unable to fold properly, chaperons prevent its aggregation and escort it for ubiquitination.

Ubiquitin is a small, highly conserved polypeptide and a major marker for degradation of damaged proteins. Proteins destined for degradation are marked by covalent attachment of ubiquitin, a multi-step, ATP dependent process referred to as ubiquitination (Figure 6). The isopeptide bond is formed between the amino group of the lysine side chain on the target protein and the carboxy-terminal glycine residue on the ubiquitin. Three main types of proteins are required for ubiquitination: ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2) and lastly ubiquitin-protein ligase (E3). Ubiquitin is synthesized as a precursor and processed by E1 enzyme to expose the carboxy-terminal glycine residue. From E1, ubiquitin is transferred to E2 enzyme. Depending on the ubiquitin ligases family type, E2, in complex with E3, can directly transfer the ubiquitin to the target protein, or, E2 first transfers the ubiquitin to the E3 enzyme, which in turn catalyzes thioester bond formation between the glycine on the ubiquitin and the active lysine on the target protein. The mode of action depends on the type of E3 ligase and the protein target, and E2 and E3 enzymes provide specificity to the system.

Following ubiquitination, the misfolded protein is targeted for degradation by the proteasome. For proteolytic degradation, ubiquitin is removed and recycled, and proteins are unfolded in order to pass through the proteasome (Figure 1). In the proteasome, proteins are cleaved into short peptides (7 - 9 amino acids long). Main shortcoming of the proteasome is - due to its constrained shape - only proteins of a limited size can be degraded by this system. Misfolded proteins and large aggregates which escape the proteasome are degraded by the second degradation system: lysosome-mediated pathway.

Autophagy is a lysosome-mediated turnover system evolutionarily conserved across eukaryotic systems. Autophagy, whose name stems from Greek and means ‘self-eating’, mediates degradation of not only large protein aggregates, but also damaged organelles and microbes. This system has many more functions than initially expected. Apart from its canonical role in starvation and degradation of bulk cytoplasmic materials, autophagy is involved in a vast range of different physiological and pathophysiological processes, summarized in Figure 2.

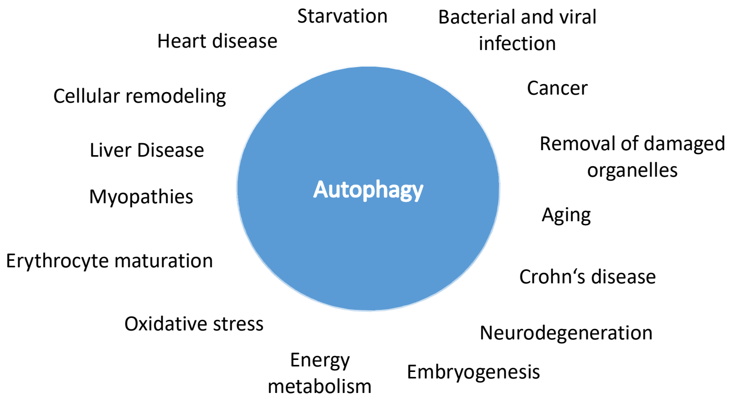


Figure 2: Autophagy and its roles. Apart from its basic role in starvation and removal of damaged organelles, autophagy is intertwined with many other cellular processes, some of which are presented in this figure (adapted from Klionsky, 2011).

1.2 Autophagy types

Autophagy can be categorized into three main types: chaperone-mediated autophagy, microautophagy and macroautophagy (Figure 3). In contrast to microautophagy and macroautophagy, which involve vesicles for delivering of cargo to the lysosome, chaperone-mediated autophagy (CMV) allows direct targeting of proteins and their delivery to the lysosomal lumen. Proteins detected by CMV harbor a specific pentapeptide KFERQ-like motif and are bound by a chaperon complex (heat shock 70 and its co-chaperones) in the cytoplasm. Chaperon complex brings the target protein to the surface of the lysosome and binds the cytosolic tail of lysosome-associated membrane protein type 2A (LAMP-2A), which in its free form is a monomer. Once bound to the target protein, LAMP-2A oligomerizes and forms a translocation pore. Upon transfer of the target protein to the lumen of the lysosome, LAMP-2A oligomer is quickly disassembled into single units, ready to bind the next target. Main inducers of CMV are prolonged starvation and oxidative stress (Cuervo *et al.* 2014).

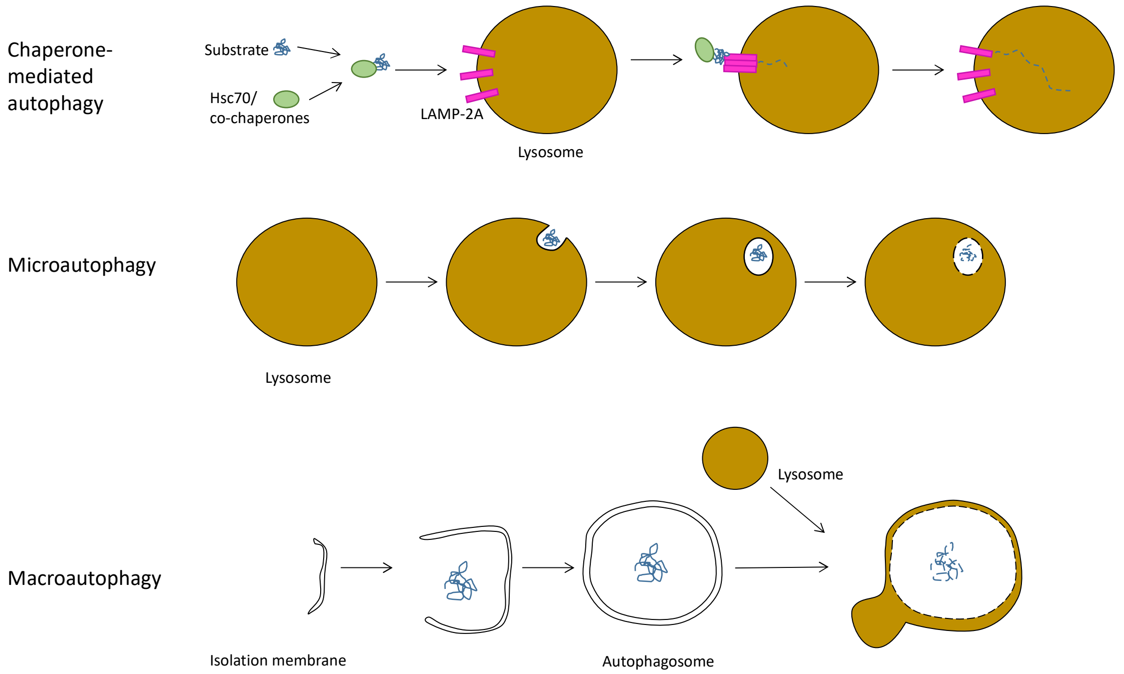


Figure 3: Types of autophagy. Chaperone-mediated autophagy is a turn-over process targeting specific signal sequence on proteins whose capture and delivery to the lysosome is facilitated by Hsp 70 and its co-chaperons. Microautophagy is characterized by formation of cup-like invaginations or finger protrusions by the lysosomes and engulfment of neighboring sections of the cytoplasm. Macroautophagy is characterized by de novo formation of a membrane which expands around cargo and fuses to form a double-membrane autophagosome. Mature autophagosome fuses with the lysosome and delivers contents for degradation. (adapted from Okamoto 2014)

Unlike CMV, which only targets proteins with the signal pentapeptide, micro- and macroautophagy can be both selective and nonselective (bulk) processes. In the case of nonselective microautophagy in mammalian systems, lysosomes have the ability to either form cup-like invaginations of their membrane, or to extend arm-like protrusions which can fold back onto itself. In both cases, the result is formation of intralysosomal vesicles containing cytosolic constituents which are degraded by lysosomal enzymes (Figure 3).

In the case of macroautophagy, the most prominent physical characteristic is formation of a double-membrane vesicle, the autophagosome. As autophagosomal membrane grows, it engulfs part of the cytoplasm. Mature autophagosomes eventually fuse with the lysosomes, delivering the contents for degradation (Figure 3). The morphology and the mechanism of action of macroautophagy are conserved from yeast to humans. So far, mouse models only exist for macroautophagy, and extensive research on this pathway has revealed clear connections between macroautophagy and human diseases. This work is focused on macroautophagy, and therefore in the sections hereafter - unless specified otherwise - macroautophagy will be referred to as autophagy.

1.3 Autophagy core machinery

Autophagy has been extensively studied in budding yeast *Saccharomyces cerevisiae*, with more than 30 proteins (encoded by autophagy-related *ATG* genes) identified as being essential for the process. More than half of the yeast autophagy proteins have homologs in mammalian systems. A subset of *ATG* genes is involved in all subtypes of autophagy and is collectively termed as the core molecular machinery. Core machinery proteins are further classified into functional groups. (Nakatogawa *et al.* 2009; Yang *et al.* 2010; Levine *et al.* 2011) These classifications are shown in Table 1.

Core subgroups	**Core mammalian proteins belonging to the subgroup**
unk-51-like kinase (ULK1) complex	ULK1, ULK2, ATG13, FIP200, ATG101
transmembrane ATG9 and associated proteins	ATG9, ATG2, WIPI1-4, DFCP1, VMP1
class III phosphatidylinositol-3-OH kinase (PI(3)K) complex	Vps34, p150 (Vps15), ATG14L, Beclin1, ultraviolet irradiation resistance associated gene (UVRAG), Ambra1, Rubicon
ubiquitin-like conjugation system LC3 and ATG12/ATG5/ATG16	ATG4, LC3/GABARAP, ATG12, ATG3, ATG7, ATG10, ATG5, ATG16

Table 1: List of core autophagy proteins in mammalian system and their functional classifications.

The process of autophagy can be roughly divided into three morphological stages: initiation, elongation, and maturation and fusion with lysosome (Figure 4).

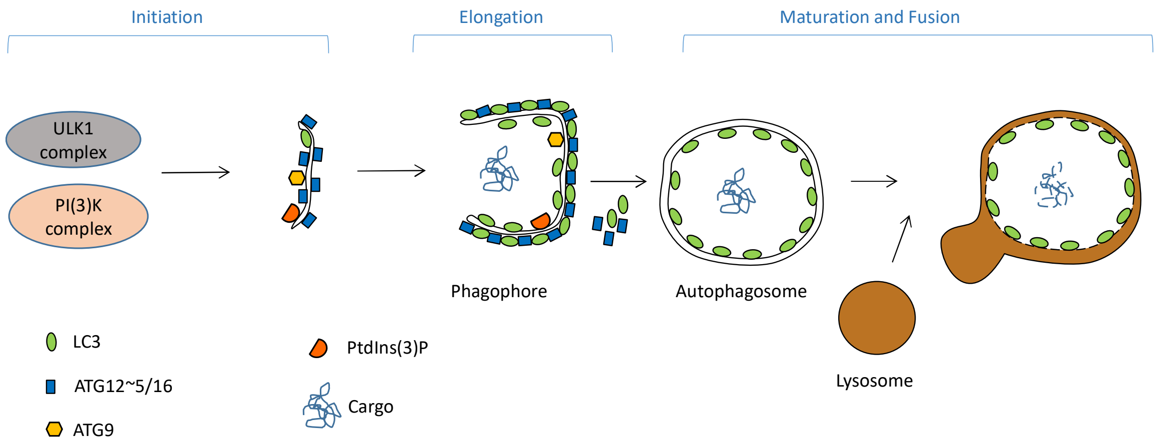


Figure 4: Overview of autophagy machinery. Autophagy can be divided into three major stages: initiation, elongation, maturation and fusion. Each step is orchestrated by a subset of core autophagy components. ULK1 complex is important for induction of autophagy and subsequent recruitment of other members of autophagy machinery. PI(3)K complex plays an essential role in nucleation. Phagophore elongation requires ATG9 and two ubiquitin-like systems (ATG12 and LC3). Each step is further described in the text.

1.3.1 Autophagosome initiation

Most of the core ATG proteins are recruited to phagophore assembly site (PAS), the site of autophagosome formation. Under nutrient rich conditions, mTOR is found in association with ULK1 complex and phosphorylates ULK1 and ATG13, keeping it in the inhibited state. Upon induction of autophagy (i.e. nutrient depletion or rapamycin treatment) mTOR is released from the complex and the ULK1 kinase is activated by partial dephosphorylation. Active ULK1 autophosphorylates and phosphorylates mammalian ATG13 and FIP200 (Figure 5) (Hosokawa *et al.* 2009; Yang *et al.* 2010; Mercer *et al.* 2009). In addition to ULK1, ATG9 appears to be one of the first proteins recruited to PAS from late endosome and Golgi membranes, where it resides under nutrient-rich conditions (Orsi *et al.* 2012; Young *et al.* 2006). However, whether ULK1 recruits ATG9 or the other way around requires further investigation (Feng *et al.* 2014; Orsi *et al.* 2012). Another equally important initiation complex is class III phosphatidylinositol-3-OH kinase (PI(3)K) complex, composed of Vps34, p150, ATG14L and Beclin-1. Ambra1 retains PI(3)K complex at the microtubules and, upon phosphorylation of Ambra1 by ULK1, the complex is released and allows translocation to the site of phagophore formation. ULK1 has also been shown to phosphorylate Beclin-1 (Russell *et al.* 2013) and thus enhance activity of the PI(3)K complex. Once both ULK1 and PI(3)K complexes are recruited to the PAS, Vps34, a member of PI(3)K complex, generates phosphatidylinositol-3-phosphate (PtdIns(3)P). PtdIns(3)P recruits autophagy effectors such as DFCP1 (double FYVE-containing protein 1) and WIPI (WD-repeat domain phosphoinositide-interacting) proteins, which recruit other ATG core proteins important for membrane elongation (Polson *et al.* 2010).

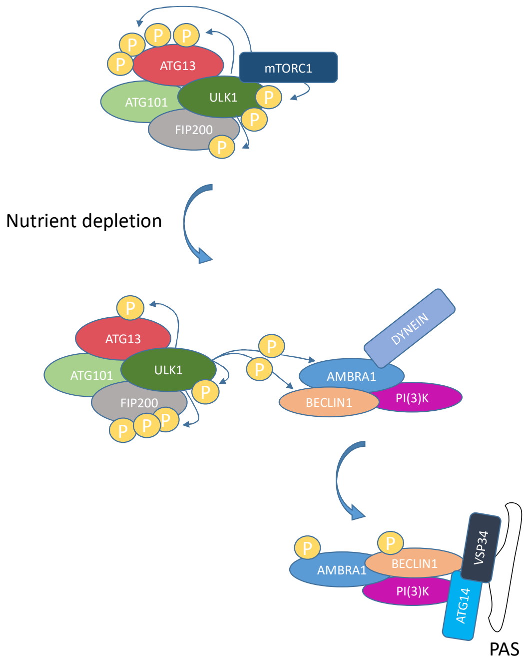


Figure 5: Core autophagy complexes involved in nucleation and initiation of autophagy. Under nutrient-rich conditions, ULK1-ATG13-FIP200 complex is kept inactive (phosphorylated) by and bound to mammalian TORC1. Once nutrient supply becomes limited, partial dephosphorylation of the ULK1-ATG13-FIP200 complex releases the complex from mTORC1. Active ULK1 phosphorylates AMBRA1 and releases it, in complex with Beclin-1 and PI(3)K from microtubules. Beclin-1 is also phosphorylated by ULK1. Vps34 is another member of the PI(3)K complex essential for generating PtdIns(3)P at the PAS.

1.3.2 Autophagosome elongation

Further growth of the autophagosomal membrane depends on two ubiquitin-like systems: ATG12~ATG5/ATG16L1 and LC3. ATG12~ATG5/ATG16L1 complex interacts with ULK1 complex (ULK1/ATG13/FIP200/ATG101) probably by direct binding of ATG16L1 to FIP200 (Gammoh *et al.* 2013; Nishimura *et al.* 2013) Once PtdIns(3)P are generated, further association of ATG16L1 with the expanding membrane is facilitated via WIPI2-ATG16L1 complex (Dooley *et al.* 2014). ATG16L1 binds ATG5~ATG12 conjugate and serves as E3 for conjugation of LC3 with phosphatidylethanolamine (PE) on the phagophore and autophagosome membrane. Both conjugation reactions (ATG5~ATG12 and LC3~PE) are mechanistically similar to that of canonical ubiquitination. Shortly, in canonical ubiquitination, precursor ubiquitin is processed by a protease which cleaves it at a glycine residue. Cleaved ubiquitin is activated by binding to E1, in an ATP dependent manner. From E1, ubiquitin is transferred to E2. Depending on the type of ubiquitin ligase, either E2 directly, or E3, after ubiquitin transfer from E2, catalyzes the final conjugation of ubiquitin to the lysine on the target protein. In autophagy, ATG12 is not previously cleaved and is directly activated by E1-like enzyme ATG7. E2-like enzyme ATG10 drives the conjugation between the active glycine on ATG12 and lysine on ATG5, resulting in covalently linked ATG12~ATG5 complex. (Mizushima *et al.* 1998) In this case, E3 enzyme is not needed, and the specificity for the substrate comes from E2. Notably, unlike ubiquitin which has many different substrates, ATG12 is exclusively conjugated to ATG5 which can bind ATG16L1, generating ATG12~ATG5/ATG16L1 complex (Mizushima *et al.* 1999) (Figure 6B).

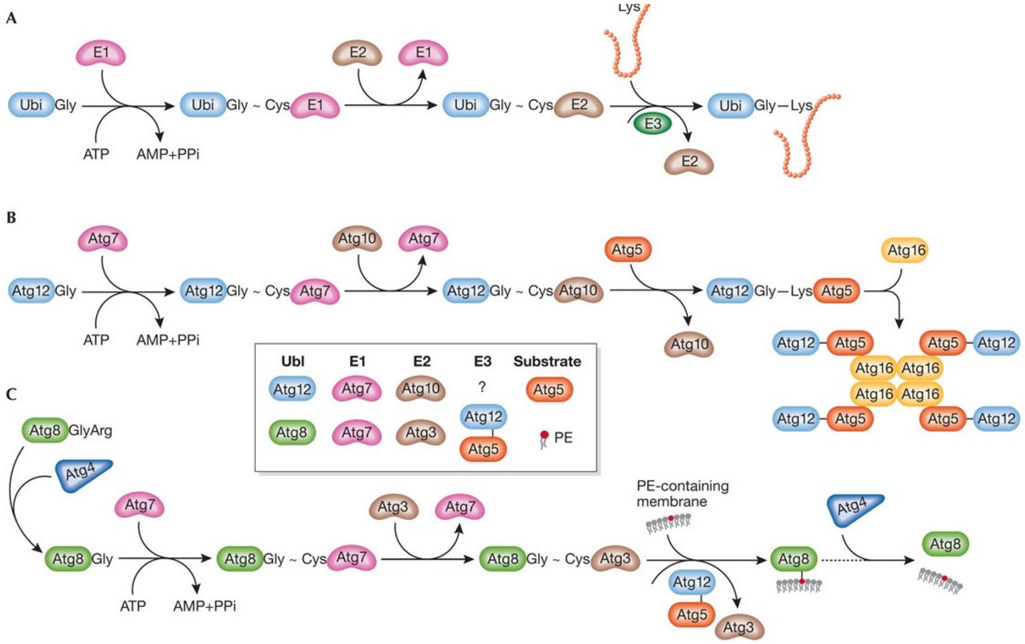


Figure 6: Ubiquitin-like conjugation systems required for autophagosome elongation. A) Canonical ubiquitination process. Ubi=ubiquitin. E1=ubiquitin activating enzyme. E2=ubiquitin conjugating enzyme. E3= ubiquitin ligase. Precursor ubiquitin molecule is activated by E1 enzyme in ATP dependent manner. Activated ubiquitin is transferred to E2 which transfers the ubiquitin to E3, which is specific for the substrate protein. E3 catalyzes final covalent link between ubiquitin and protein, marking the protein for degradation. B) ATG12 ubiquitin-like conjugation system. ATG7 is E1-like enzyme. ATG10 is E2-like enzyme. ATG12 is conjugated to ATG5. ATG12~ATG5 conjugate binds ATG16 via ATG5 C) LC3 (mammalian homolog of yeast ATG8) ubiquitin like system. LC3/ATG8 is conjugated to the lipid (PE) on the phagophore and autophagosome, instead to another protein. LC3/ATG8 precursor is cleaved by ATG4. ATG7 is E1-like enzyme. ATG3 is E2-like enzyme. Role of E3 enzyme is played by ATG12~ATG5/ATG16. LC3/ATG8 is bound to PE on the membrane and can be cleaved of by the ATG4. (Geng et al. 2008)

Studies of yeast homologues predicted this complex to form a tetramer of ATG12~ATG5/ATG16 (Fujioka *et al.* 2010). However, recent studies of the coiled-coil domain of human ATG16L1 show it to form a dimer in solution (Parkhouse *et al.* 2013). ATG16L1 dimerizes through a coiled-coil domain in the middle of the protein resulting in a complex likely to be composed of two copies of ATG12~ATG5/ATG16L1.

In the case of the second ubiquitin-like system in autophagy, yeast ATG8 is cleaved by a cysteine protease ATG4 which exposes the glycine, prior to activation by the E1-like enzyme ATG7. In this case, E2-like enzyme ATG3 takes over the ATG8. (Ichimura *et al.* 2000) Final covalent link between ATG8 and phosphatidylethanolamine (PE) on the phagophore and autophagosome membrane is catalyzed by ATG12~ATG5/ATG16 complex, an unusual E3-like enzyme (Fujita *et al.* 2008) (Figure 6C). The exact mechanism of this process is still under investigation (Geng *et al.* 2008). Interestingly, unlike canonical ubiquitination and ATG5~ATG12/ATG16, ATG8 is not conjugated to another protein, but to a lipid on an autophagosome membrane. Mammalian homologs of the ATG8 are divided into two main subgroups, based on their sequence homology: LC3 (LC3A, LC3B, LC3C) and GABARAP (GABARAP, GABARAP L1, GABARAP L2 or GATE16, GAABRAPL3) which all undergo similar modifications described for the yeast ATG8 (Kabeya *et al.* 2004; Kabeya *et al.* 2000). Another mammalian homologue of yeast ATG8 has been reported (ATG8L) (Hemelaar *et al.* 2003) and shown to be a substrate of ATG4, ATG7, and ATG3 (Tanida *et al.* 2006). Due to the tight association between LC3 and PE, the conversion between unbound, cytosolic LC3 (LC3I, 18kDa) and LC3-PE (LC3II, 16kDa), LC3 has been best characterized as a cellular readout of autophagy levels and is commonly detected by Western blot to monitor autophagy. (Mizushima *et al.* 2014)

Both ubiquitin-like systems are essential for an expansion of the autophagosome. Apart from acting like an E3 enzyme, studies using the yeast homologues showed that ATG12~ATG5/ATG16 complex also marks the exact location of ATG8 lipidation (Suzuki *et al.* 2001; Fujita *et al.* 2008). ATG12~ATG5/ATG16 complex primarily localizes on the outer membrane, while ATG8 is present on both, outer and inner face of the membrane of the growing autophagosome. Shortly before or upon completion of the autophagosome, ATG12~ATG5/ATG16 complex dissociates from the outer membrane (Figure 4) (Mizushima 2003; Mizushima *et al.* 2001). Lipidated LC3 also dissociates from the outer autophagosomal membrane prior to autophagosome/lysosome fusion (Yang *et al.* 2010).

1.3.3 Autophagosome maturation and lysosomal fusion

Fusion event has been extensively studied, and autophagosomes have been observed to fuse with early and late endosomes, and with lysosomes (Dunn 1990; Eskelinen 2005). TECPR1 (tectonin β-propeller repeat-containing protein 1) a lysosomal membrane protein has been identified as a component of autophagy network (Behrends *et al.* 2010). TECPR1 exists in an auto-inhibited form which is disrupted by its binding to ATG5 (of the ATG12~ATG5/ATG16L1 complex). TECPR1 displaces ATG16L1 from the ATG5~ATG12 conjugate. Once in complex with the ATG5~ATG12, phosphoinositides-binding domain of TECPR1 becomes available to interact with PtdIns(3)P on the autophagosomal membrane, thus providing specificity of the fusion event. (Chen *et al.* 2012) Autophagosome maturation and its fusion with the lysosome are tightly regulated events. Many additional proteins involved in the final stages of autophagy have been identified so far, including: LC3/GABARAP proteins (mammalian homologues of ATG8) (Wang *et al.* 2015), class III phosphatidylinositol 3-kinase and Rubicon (Sun *et al.* 2011), SNAREs (Moreau *et al.* 2011; Nair *et al.* 2011), and a small GTP binding protein Rab7 (Jäger *et al.* 2004). However, the exact mechanism of the fusion events is still not clear.

1.4 Selective autophagy

Autophagy has long been considered an essential but non-selective bulk degradation pathway. However, mounting evidence suggests the selective elimination of atypical protein aggregates, dysfunctional organelles and invading pathogens by autophagy. Based on the remarkable substrate specificity of this process, selective autophagy is classified into several types, including: mitophagy (selective degradation of mitochondrion), nucleophagy (selective degradation of parts of nucleus), pexophagy (selective degradation of peroxisomes) and xenophagy (selective degradation of intracellular pathogens) (Figure 7).

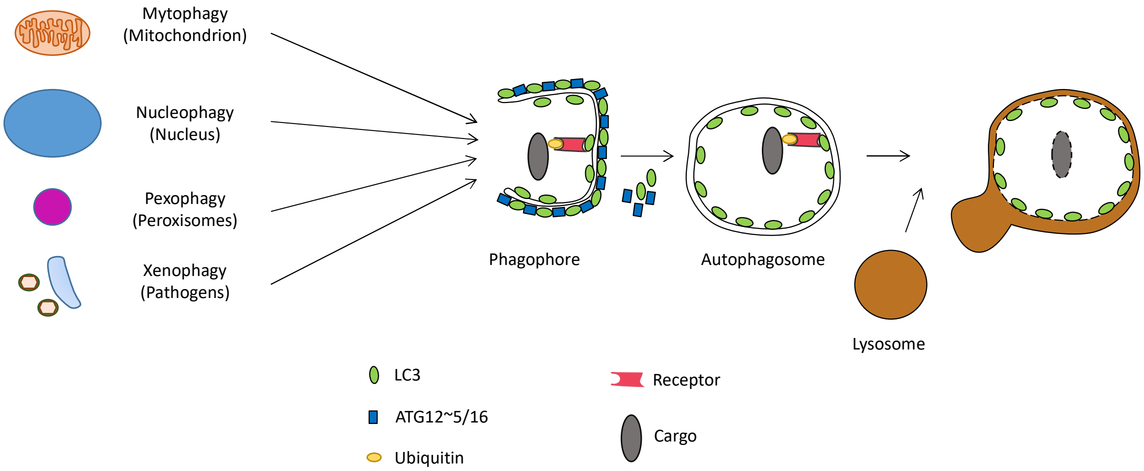


Figure 7: Modes of selective autophagy. Damaged or dysfunctional organelles (mitochondrion, nucleus and peroxisome) as well as intracellular pathogens (including pathogen-containing phagosomes and their remains) are tethered to the autophagosome by a specific molecule which binds ubiquitin on one side, and LC3/GABARAP on the autophagosome membrane on the other. In this way, autophagy receptors selectively deliver cargo for degradation by autophagy.

Specific cargo is targeted for degradation by specific autophagy receptors and adapters. The discovery and characterization of cargo-recognizing receptors revealed some mechanistic details of selective autophagy. Autophagy specific receptors are dispensable for bulk, non-specific autophagy. These molecular receptors recognize and bind ubiquitinated substrates, tether them to the site of autophagosome engulfment, and act as a physical link between substrates and autophagosome-specific proteins, such as the members of the LC3/GABARAP family (Johansen *et al.* 2014; Isakson *et al.* 2013). The binding between autophagy receptors and LC3/GABARAP family members is facilitated through a short hydrophobic sequence (D/E-D/E-D/E-W/F/Y-X-X-L/I/V) on the receptors, termed LIR (LC3-interacting region) motif (Bjørkøy *et al.* 2005; Johansen *et al.* 2014). The growing list of autophagy receptors includes: p62 (Bjørkøy *et al.* 2005), NBR1 (neighbor of Brca1 gene) (Kirkin *et al.* 2009), NDP52 (Thurston *et al.* 2009) and optineurin (Wild *et al.* 2011). These autophagy receptors have an ubiquitin-recognition domain and at least one LIR domain. In this way, they can bind ubiquitinated cargo on one end and LC3 on the autophagosome membrane on the other. p62 plays a role in mitophagy (Ding *et al.* 2010; Geisler *et al.* 2010), xenophagy (Dupont *et al.* 2009; Yoshikawa *et al.* 2009; Zheng *et al.* 2009) and pexophagy (Kim *et al.* 2008); while NDP52 and optineurin have so far been linked only to xenophagy (Thurston *et al.* 2009; Wild *et al.* 2011). So far, autophagy receptor p62 has been shown mediate xenophagy for *Salmonella enterica* serovar Typhimurium *(S.* Typhimurium*)* (Zheng *et al.* 2009), *Shigella flexneri* (Dupont *et al.* 2009), and *Listeria monocytogenes* (Yoshikawa *et al.* 2009). In addition to p62, *S.* Typhimurium is restricted by optineurin and NDP52 (Wild *et al.* 2011). The molecular machinery of selective autophagy is just starting to be unraveled, and further work is needed to deeper explore which autophagy receptors are involved in recognition of the various types of cargo for selective autophagy.

1.4.1 Autophagy in pathogen defense

One kind of selective autophagy, xenophagy, eliminates intracellular pathogens by restricting their replication and survival. Xenophagy is able to target damaged and intact pathogen-containing vacuoles as well as free pathogens in the cytosol and transport them to the lysosome for degradation (Figure 8).

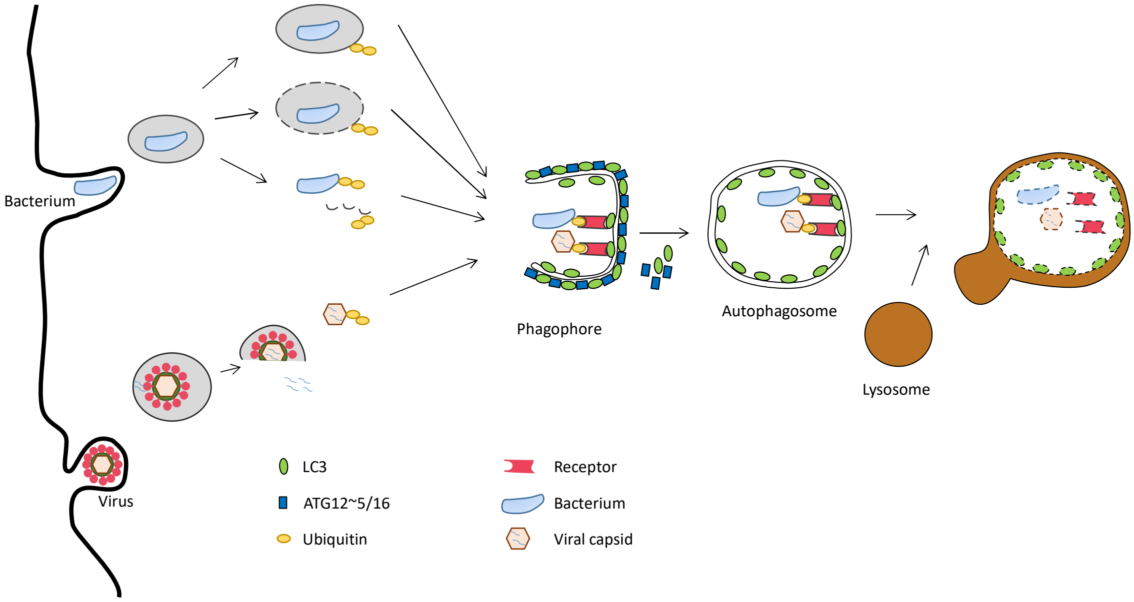


Figure 8: Xenophagy. Intracellular viruses and bacteria, cytosolic and vacuolar, are recognized and ubiquitinated by an unknown mechanism, marking them for degradation by autophagy. Ubiquitinated pathogens recruit autophagy receptors (note: the receptor is colored pink in the figure for simplicity purpose, however there is more than one receptor involved in xenophagy), which tether them to the expanding phagophore where they are engulfed by autophagosome and delivered to the lysosome for degradation.

In the case of *Salmonella* Typhimurium, autophagy contained the infection at the early stage by targeting damaged Salmonella-containing-vacuoles and prevented colonization of the cytoplasm (Birmingham *et al.* 2006). *Streptococcus pyogenes* escapes the phagosomes upon invasion of nonpathogenic cells. Once free in the cytosol, *S. pyogenes* is captured in the autophagosomes and delivered to the lysosomes for degradation. In this way, autophagy effectively restricted pathogenic growth. (Nakagawa *et al.* 2004) *Mycobacterium tuberculosis* is an intracellular pathogen which resides inside phagosomes and prevents their maturation. Autophagy induced by starvation increased localization of Beclin-1 with mycobacterium-containing phagosomes. Accumulation of critical autophagy proteins eventually overrides the block of phagosomal maturation and inhibits mycobacterial survival. (Gutierrez *et al.* 2004) Autophagy is not specific against bacterial invasions. *Toxoplasma gondii* is an obligate intracellular protozoan. Once inside the cell, *T. gondii* resides within a vacuole void of any proteins which could recruit lysosomes. Treatment of *T. gondii* infected macrophages stimulated fusion of *T. gondii* containing vacuoles with lysosomes through autophagy. (Andrade *et al.* 2006) Also viruses have been reported to succumb to the process of autophagy (Tallóczy *et al.* 2005; Orvedahl *et al.* 2010a) Based on these and similar studies, autophagy has emerged as a specialized cytoplasmic system for elimination of pathogens. However, autophagy is not a foolproof mechanism in pathogenic defense. Many pathogens have developed ways to escape autophagy or to take advantage of the autophagy machinery for their benefit (Ogawa *et al.* 2005; Pizarro-Cerda *et al.* 1998; Beron *et al.* 2002; Dorn *et al.* 2001).

Interplay between autophagy and immunity does not stop at clearance of cytosolic pathogens. Once pathogen-containing autophagosome fuses with the late endosome, viral, bacterial and parasitic antigens are presented on MHC (major histocompatibility complex) class II molecules for priming of T ($CD4^+$) cell specific response (Schmid & Münz 2007).

Two main open questions remain about xenophagy (and selective autophagy in general): how are substrates recognized and sorted into the autophagy pathway? Which ubiquitin ligases are important for marking the substrates for degradation by autophagy?

1.4.1.1 ATG16L1: link between autophagy and pathogen control?

To better understand the connection between autophagy and immunity two main questions have to be answered. One question is how pathogens become labelled for degradation by autophagy. Due to a vast spectrum of pathogens targeting human cells, it is not easy to answer this question. Some evidence comes from studies with *S.* Typhimurium. Membranes of Salmonella-containing vacuoles become damaged by bacteria entering the cytosol. Galectin - 8 is recruited as a danger molecule to damaged membranes and its presence is specifically sensed by NDP52 (Thurston *et al.* 2012). Galectin - 8 also coats phagocytes harboring *S. flexneri* and *L. monocytogenes*, indicating that also these membranes, once damaged by escaping bacteria, are sensed though galectin - 8/NDP52 link. Once free in the cytosol, bacteria are poly-ubiquitinated (Perrin *et al.* 2004). Based on this, both galectin - 8 and ubiquitin serve as markers for uptake by autophagy, and both are recognized by NDP52. More extensive research is needed to confirm this mechanism for other intracellular pathogens or to reveal novel marker molecules.

Since ubiquitination is a factor in marking pathogens for autophagy, ubiquitin ligase responsible for labeling cytosol-exposed bacteria for autophagy degradation would provide a way to assess the importance of the ubiquitin coat in the antipathogenic defense. As a candidate for pathogen-specific ubiquitin ligase, DDB1-CUL4 complex emerged from the study performed by He *et al.* (He *et al.* 2006).

1.4.1.2 ATG16L1: a true DCAF?

DDB1-CUL4 is a member of a large cullin-RING E3 ubiquitin ligase family. Binding of RING family members to E2 is either mediated by an intrinsic RING finger domain or a small RING protein, such as ROC1/RBX1. There are at least seven cullins, and each can assemble into specific cullin-RING ligase complexes (Lee *et al.* 2007). The cullin complexes are highly modular and each of the cullins binds a specific adaptor or adaptor/receptor molecule for recruitment of distinct substrates for ubiquitination (Figure 9A). The above mentioned CUL4, binds ROC1/RBX1 at the C-terminal end while the other side of the arc-shaped protein recruits its adapter, DDB1 (DNA damage binding protein 1) forming the DDB1-CUL4 ubiquitin ligase complex. In turn, the adaptor DDB1 recruits receptor molecules, a wide range of WD40 domain containing proteins, collectively termed DDB1-CUL4 associated factors (DCAFs). These DCAFs act as binding platform for the ubiquitination substrate and provide substrate specificity to the DDB1-CUL4 ubiquitin ligase complex. Sequentially, the adaptor DDB1 recruits receptor molecules.

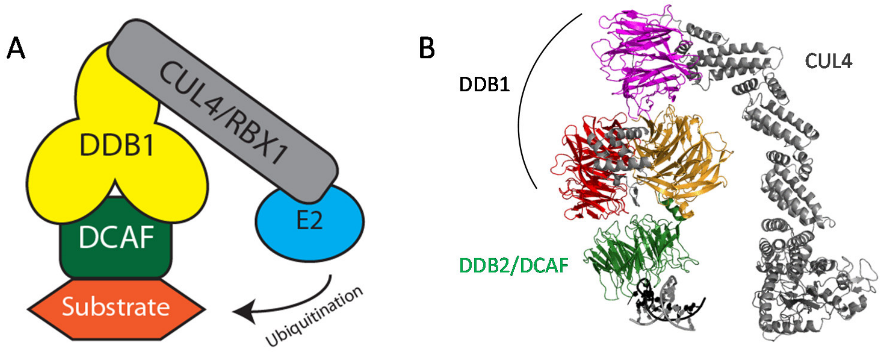


Figure 9: Graphical and structural model of CUL4-DDB1-DCAF. A) Graphical representation of CUL4-DDB1-DCAF mode of action. DCAF recruits a substrate which is ubiquitinated by E2 enzyme bound to the C-terminal end of CUL4. B) Structural model of CUL4-DDB1 bound to DCAF (DDB2) and substrate (DNA). Arc-like structure of CUL4 is represented in gray. DDB1 domains are separately colored: β-propeller B responsible for binding to CUL4 is in purple. The other two domains (β-propeller A in red and C in yellow) form a cleft for DCAF binding (here, DDB2 in green). DDB2 recruits DNA, represented as part of a-helix in gray, allowing for ubiquitination of DNA-associated proteins (Scrima *et al.* 2011)

Structure of DDB1, in complex with CUL4 and DDB2 (a DCAF involved in DNA damage repair), is shown in Figure 9B and reveals three WD40 propeller domains of DDB1 (Fischer & Scrima *et al. 2011*). One of the domains (BPB) is involved in binding to CUL4, while the other two form a pocket for binding of the DCAF. (Angers *et al.* 2006; He *et al.* 2006) He *et al.* identified a member of core autophagy machinery, ATG16L1, as a DCAF, suggesting a role for ATG16L1 as receptor molecule involved in recruitment of ubiquitination substrates for DDB1-CUL4 ubiquitin ligase complex. Furthermore, they identified a DWD (DDB1-binding and WD40 repeat) box, a 16 residue motif recognized by CUL4-DDB1. Alanine substitution of a conserved arginine in the DWD box on ATG16L1 abolished the DCAF phenotype (He *et al.* 2006). Structural and biophysical studies are needed to confirm the direct interaction between ATG16L1 and DDB1, and the importance of DWD box for this interaction.

1.4.1.3 ATG16L1: a link between pathogen sensing and pathogen degradation?

If ATG16L1 is a true DCAF and thus recruits substrates (pathogens) for ubiquitination, the question still remains: how are pathogens recognized as targets of autophagy? Literature search provides multiple accounts linking ATG16L1 to pattern recognition receptors (PRRs). PRRs recognize various pathogen-associated molecular patterns (PAMPs) and initiate a cascade of events to fight the infection. PRRs can be membrane bound, such as Toll-like receptors (TLRs) or cytosolic, such as nucleotide-binding oligomerization domain (NOD) receptors. (Kawai *et al.* 2006) First characterized members of NOD family, NOD1 and NOD2, localize to the plasma membrane, sense components of peptidoglycan bacterial wall, and initiate a proinflammatory response (Girardin *et al.* 2003). NOD1 senses a unique diaminopimelate-containing GlcNAc-MurNAc tripeptide muropeptide (GM-Tri$_{DAP}$) present only on the wall of Gram-negative bacteria (Girardin *et al.* 2003; Chamaillard *et al.* 2003), while NOD2 is a sensor for muramyl dipeptide (MDP, N-acetylmuramyl-L-analyl-D-isoglutamine), a more general, minimal active motif on both Gram-positive and Gram-negative peptidoglycan wall (Inohara *et al.* 2003; Girardin *et al.* 2003).

A study done by Travassos *et al.* in 2010 reported that mice, which were induced with NOD1 and NOD2 agonists, showed a robust increase of autophagy in peritoneal macrophages. Additionally, they demonstrated that NOD1 and NOD2 co-localize with ATG16L1 on the plasma membrane, surrounding the invading bacteria (*Shigella flexneri*). ATG16L1 lacking the first 85 amino acids (required for interaction with ATG5) exhibited the same behavior as full length ATG16L1. The interaction was confirmed with endogenous ATG16L1 and NOD proteins, as well as with co-immunoprecipitation. (Travassos *et al.* 2010; Ramjeet *et al.* 2010) Furthermore, NOD1 and NOD2 induced by bacterial entry recruit ATG16L1 and induce autophagy in a NF-κB independent manner (Travassos *et al.* 2010), unlike previously demonstrated mode of action (Abbott *et al.* 2004; Inohara *et al.* 2000). Based on their findings, Travassos *et al.* proposed a mechanism, illustrated in Figure 10, by which NOD2 recruits ATG16L1 to the site of the invading bacteria. By recruiting ATG16L1, entire autophagy machinery is brought into close proximity with the pathogen, which is engulfed by autophagosome and delivered to the lysosome for degradation (Travassos *et al.* 2010).

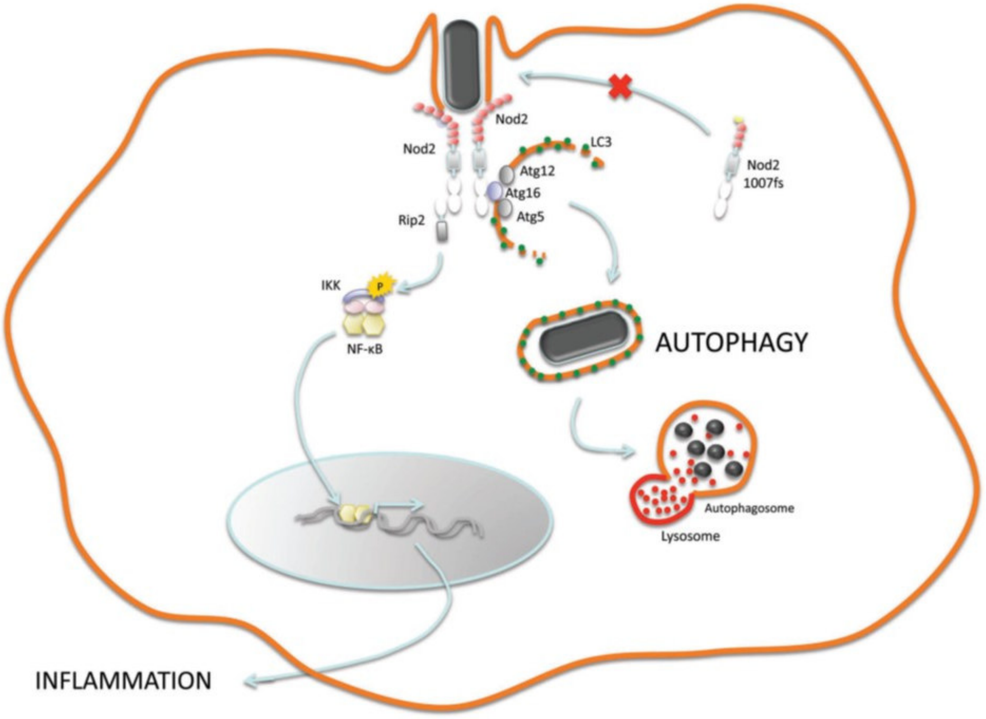


Figure 10: Proposed mechanism of autophagy recruitment to the bacterial entry site by NOD2. NOD2 is a cytosolic pathogen recognition receptor which is recruited to the plasma membrane by invading pathogens and senses muramyl dipeptide on the bacterial peptidoglycan wall via leucine-rich domain (pink circles). Upon activation, NOD2 triggers two pathways: proinflammatory response via NF-κB pathway, and autophagy by recruiting ATG16L1. Variant of NOD2 (L1007fs) most commonly present in Crohn's disease patients is not recruited to the plasma membrane upon bacterial invasion and fails to activate either pathway. (Travassos et al. 2010)

What further strengthens the NOD2-ATG16L1 link between pathogen sensing and autophagy is the presence of mutations on both proteins in patients suffering from Crohn's disease (CD). CD is a chronic inflammatory disease of the gastrointestinal (GI) tract. Any part from mouth to anus can be affected, but CD is most commonly observed in the ileum. Depending on the severity of the disease, it can manifest itself as small erosions ranging to deep ulcers of the inner surface of the GI wall. Deep and large ulcers can lead to obstruction of the bowel and formation of physical holes in the wall of the GI tract. In such severe cases, disease can spread to the neighboring organs. Approximately half a million people are affected by CD in North America and Europe (Naser *et al.* 2012). CD is difficult to diagnose since it can affect any part of the GI tract, and its symptoms (diarrhea, abdominal pain, fever and weight loss) can be easily associated with many other conditions. The cause of CD is not known in detail but is attributed to a combination of environmental factors, genetics, immune dysfunction and microbiota composition (Sartor 2005). There is no cure for CD. Treatment of symptoms depends on the severity of the disease and includes nutritional management combined with therapeutic drugs. In most severe cases, surgery and removal of damaged areas are considered (Beattie *et al.* 2006). Based on genetic screens and population studies, *NOD2* was the first identified CD susceptibility gene. Two missense mutations (R702W and G908R) and one frameshift mutation (L1007fsinsC (L1007fs)) on NOD2 have been associated with increased predisposition to CD (Hugot *et al.* 2001). Population studies have identified L1007fs variant of NOD2 as prevalent in CD patents on European and North American continents (Hampe *et al.* 2001; Ogura *et al.* 2001; Sventoraityte *et al.* 2010; Vermeire *et al.* 2002; Hugot *et al.* 2001). While studying the link between NOD2 and ATG16L1, Travassos *et al.* also revealed that L1007fs variant of NOD2 does not get recruited to the site of bacterial entry, nor does it induce NF - κB signaling, nor does it recruit ATG16L1 and autophagy (Figure 10) (Travassos *et al.* 2010).

Genome wide analysis of patients exposed a CD susceptibility variant in ATG16L1 (Libioulle *et al.* 2007; Hampe *et al.* 2006). A single nucleotide polymorphism in *ATG16L1* gene results in amino acid substitution of threonine with alanine at position 300 (T300A) in the amino acid sequence of the protein. This variant was shown to have the same expression level as the wild type and to be a risk factor for CD in the absence of NOD2 risk variants (Hampe *et al.* 2006). The effects this mutation has on behavior of ATG16L1 and how it could contribute to CD was considered and examined during the course of this work. In 2014, studies done in mice explained T300A variant of ATG16L1 is more prone to degradation by caspase 3 (Murthy *et al.* 2014) and caspase 7 (Lassen *et al.* 2014) resulting in protein instability and a reduction in pathogen control via autophagosome/lysosome pathway allowing bacteria to dominate (Lassen *et al.* 2014; Murthy *et al.* 2014). Moreover, NOD2 mediated autophagy leads to localization MHC II antigen presentation machinery members with LC3 and increased surface presentation of antigens on MHC II molecules. Dendritic cells harboring NOD2 L1007fs and ATG16L1 T300A failed to demonstrate same behavior. (Cooney *et al.* 2010) In addition to cellular defects already mentioned, inability to present antigens and prime CD4^+ T cell response, would lead to a persistent inflammatory response characteristic of the CD. Needless to say, NOD2 and ATG16L1 are not the only genetic risk factors associated with CD. With next generation sequencing, more and more CD patients' genomes can be analyzed and new loci are added to the growing list of variants contributing to CD (Liu & Anderson 2014). Interestingly, another autophagy protein, IRGM (immunity-related GTPase family M), contributes to the susceptibility to CD (Parkes *et al.* 2007). IRGM has been shown to play an essential role in autophagy-dependent pathogen clearance (Singh *et al.* 2006). Unlike other autophagy proteins, IRGM has no homologs in yeast, and it has been suggested that IRGM contributes to autophagy indirectly (Singh *et al.* 2010; Singh *et al.* 2006). Recently, IRGM has been shown to link members of autophagy-initiation complex (ULK1, Beclin-1, and AMBRA1) with NOD2. Moreover, IRGM co-localizes with NOD2 and ATG16L1 in co-transfected cells, through direct interaction between IRGM and NOD2 (Chauhan 2015).

These findings place three genetic risk factors for CD in a three partite complex important for bridging microbial sensing with autophagy degradation. This clearly demonstrates the complexity and importance of autophagy involvement in pathogen recognition, their control and clearance, and grave consequences defects in autophagy can have on the human system.

Connection between NOD2 and ATG16L1 would provide only part of the answer: how cytosolic pathogens get recruited to autophagosomes. While the cytosol is under NOD2 surveillance, cellular membranes are protected by members of Toll-like receptors (TLRs), a major family of membrane-bound PRRs. Among the TLRs, TLR1/TLR2 and TLR2/TLR6 heterodimers, TLR3, TLR4, TLR7/8 and TLR9 have all been shown to induce autophagy-mediated defense (Chaturvedi *et al.* 2008; Delgado *et al.* 2008; Shin *et al.* 2010). TLR1/TLR2 and TLR2/TLR6 heterodimers and TLR4 are localized on the mammalian cell surface, while TLR3, TLR7/8 and TLR9 oversee the inside of the endosomes. Involvement of TLR2 in autophagy is among the better understood. In 2007 Sanjuan *et al.* reported TLR2 signaling to recruit LC3 to the phagosomes, resulting in a fusion between LC3-positive phagosome with the lysosome (Sanjuan *et al.* 2007). Further evidence for TLR2 involvement in autophagy came in 2013, when Boada-Romero *et al.* identified a novel 19 residue ATG16L1 binding motif ([YX]-X_3-[ED]-X_4-[YWF]-X_2-L) present on intracellular part of TLR2, N-terminal CARD (CARD1) of NOD2, and on a cytosolic part of transmembrane protein TMEM59. TMEM59 is present on late endosomes/lysosomes and was shown to co-localize with LC3, ATG16L1 and intracellular *Staphylococcus aureus*. The ATG16L1 binding motif was sufficient for each of the above mentioned interaction partners to co-localize and co-precipitate with C-terminal domain of ATG16L1. (Boada-Romero *et al.* 2013) CARD domain of NOD1 was shown to bind to ATG16L1 (Travassos *et al.* 2010) but does not contain the complete ATG16L1 binding motif identified by Boada-Romero *et al.* and also does not co-precipitate with ATG16L1 (Boada-Romero *et al.* 2013). This suggests that not all ATG16L1 interaction partners bind exclusively through this motif, and further studies are needed to fully describe the repertoire of macromolecules recruited by ATG16L1. Alternate functions of the identified ATG16L1 binding motif cannot be excluded. Additional biophysical experiments and structural data is needed to endorse these interactions and confirm the relevance of the identified motif.

1.5 Structural knowledge of human ATG16L1

ATG16L1 is a member of core autophagy machinery. In bulk autophagy, during autophagosome formation, ATG16L1 recruits ATG12~ATG5 conjugate, via ATG5 interaction, to the site of LC3 lipidation on the autophagosome membrane (Fujita *et al.* 2008). ATG16L1 is a 607 amino acid long protein (~70 kDa) roughly divided into three major domains: ATG5 binding domain (ATG5-BD), coiled coil domain (CCD) and predicted WD40 domain (Figure 11A). ATG5 BD is a short helix on the N-terminal part of the ATG16L1, also present in the yeast homolog. Structure of the ATG5 BD was recently solved, in complex with ATG12~ATG5 conjugate (Figure 11B) (Otomo *et al.* 2013). Another similarity between ATG16L1 and its yeast counterpart is a coiled-coil domain (CCD), important for oligomerization of ATG16L1 molecules. So far, only CCD of the yeast ATG16 has a structural model (Fujioka *et al.* 2010). From the structure, shown in Figure 11B, ATG16 forms a parallel dimer through the CCD. In addition to being important for self-oligomerization of ATG16L1, CCD in human ATG16L1 is also a docking site for WIPI2 (Dooley *et al.* 2014) and Rab33, a small Golgi-resident GTPase (Itoh *et al.* 2008). Both WIPI2 and Rab33 serve as modulators of autophagosome elongation through interaction with ATG16L1. Additionally, FIP200, a member of the ULK autophagy initiation complex, binds ATG16L1 in the region between CCD and predicted WD-40 domain (230-300 region) (Gammoh *et al.* 2013). Dimerization domain, WIPI2-, Rab33- and FIP200- interacting domains are all independent of each other (Dooley *et al.* 2014; Gammoh *et al.* 2013; Itoh *et al.* 2008). The stretch between CCD and predicted WD-40 domain of ATG16L1 also harbors a position of a natural variant of ATG16L1 (T300A), a risk factor for Crohn's disease (Hampe *et al.* 2006) and cleavage site by caspase 3 (Murthy *et al.* 2014) and caspase 7 (Lassen *et al.* 2014).

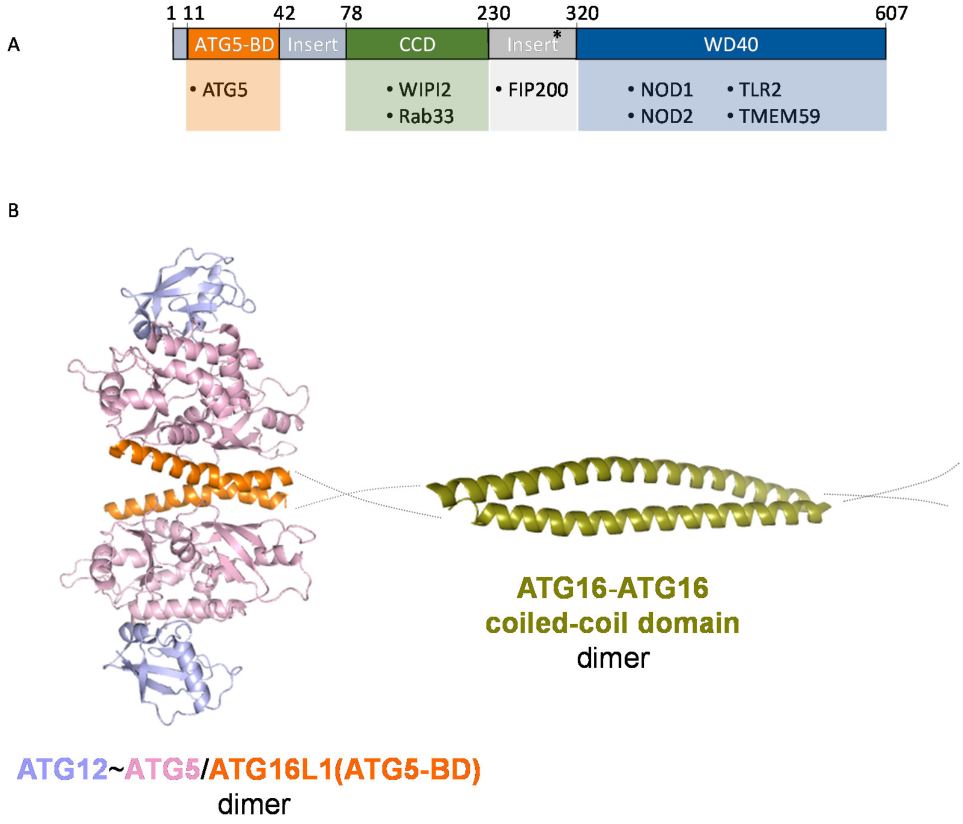


Figure 11: Current structural knowledge of ATG16L1 A) Graphical domain representation of ATG16L1: ATG5 binding domain (ATG5BD, 11-42, in orange), coiled coil domain (CCD, 78-230, in green) and WD-40 domain (320-607, in blue). Domain boundaries used here are from the UniProt (ID: Q676U5). Below each domain are interaction partners boxed in a color corresponding to the region of ATG16L1 with which they interact. An asterisk (*) indicates an approximate position of T300A mutation. B) Current structural knowledge of human and yeast ATG16. Structure of human ATG16L1 ATG5 binding domain (in orange) has been solved in complex with ATG5~ATG12 conjugate (ATG5 in purple, ATG12 in light blue) (Otomo et al., 2013)*. The structure is here modeled into a dimer using PyMOL. Structure of the coiled-coil domain of yeast ATG16 (in green) was solved and showed that the CCD domain assembles into a parallel dimer* (Fujioka et al., 2010) *indicating that ATG16 self-dimerizes.*

ATG5BD and CCD are common to both yeast and human ATG16. The major difference to the yeast protein is a predicted WD-40 propeller domain on the C-terminal part, which is exclusively present in mammalian ATG16L1. So far, no structural information is available for this part of ATG16L1. The absence of this domain from the yeast protein indicates it is not essential for bulk autophagy, but might instead have evolved to fulfill specific needs of higher eukaryotic systems, such as control and clearance of pathogens. To further support this line of reasoning, molecules mentioned so far in the concept of linking autophagy with pathogen clearance (i.e. NOD2, NOD1, TLR2 and TMEM59) have all been shown to bind to precisely this domain of ATG16L1 (Travassos *et al.* 2010; Boada-Romero *et al.* 2013).

Given the ever growing importance selective autophagy has in the mammalian system, ATG16L1, and more specifically the predicted WD-40 domain of ATG16L1, seems to provide a direct link between autophagy and other cellular pathways. WD-40 domain of ATG16L1 might be an interaction platform for various proteins while the N-terminal part of ATG16L1 remains free for interactions with ATG5 and other core autophagy proteins. However, structural details regarding the predicted WD-40 domain are needed to confirm the fold of this domain. Additionally, structural knowledge would contribute to the better understanding of its function and the overall architecture of mammalian ATG16L1.

2 Aims of this work

Autophagy is traditionally considered a bulk, unspecific, catabolic eukaryotic mechanism responsible for cellular maintenance and survival under starvation conditions. In recent decades, autophagy is being implicated in many other cellular processes, including (and not limited to) pathogenicity and immunity. Many bacteria, viruses and parasites are degraded via autophagy/lysosome pathway (Andrade *et al.* 2006; Birmingham *et al.* 2006; Gutierrez *et al*. 2004; Nakagawa *et al.* 2004; Orvedahl *et al.* 2010). However, the molecular details on pathogen recognition and recruitment by autophagy are still poorly understood.

Invading pathogens are marked by ubiquitin and degraded by autophagy. Evidence suggests that pathogenic ubiquitination is catalyzed by CUL4-DDB1, a member of a large family of ubiquitin ligases, since one of the core autophagy proteins, ATG16L1, has been proposed as a DCAF (DDB1 CUL4 associated factor) (He *et al.* 2006) and could recruit pathogens for ubiquitination. In this way, ubiquitinated pathogens can be directly engulfed by autophagy and degraded. Furthermore, ATG16L1 has been shown to bind NOD2. NOD2 is a pathogen recognition receptor which recognizes muramyl dipeptide on bacterial peptidoglycan wall and induces host's inflammatory response. In parallel, NOD2 recruits ATG16L1 to the site of invasion, and with it the rest of the autophagy machinery. (Travassos *et al.* 2010; Ramjeet *et al.* 2010) The connection between ATG16L1 and NOD2 in pathogen recognition and control is supported by the fact that mutations in both genes have been identified as risk factors for Crohn's disease, an inflammatory disorder of the gastrointestinal tract.

Taken together, this evidence identifies an autophagy core protein ATG16L1 as a good candidate for linking autophagy with pathogen control. The molecular mechanism of how ATG16L1 functions in these new roles is not very clear. A major addition to the human ATG16L1, compared to its yeast homolog, is a large (~30kDa) domain on the C-terminal end. The C-terminal half of ATG16L1 has been predicted to be a WD-40 propeller domain. Family of WD-40 propeller domains is very common in eukaryotic systems and is involved in interactions in many different cellular processes (Stirnimann *et al.* 2010). The WD-40 propeller of ATG16L1 has been demonstrated to bind both with DDB1 and NOD2, independent of the N-terminal half of ATG16L1.

The main purpose of this work is structural and biochemical characterization of WD-40 propeller of ATG16L1. Since WD-40 domains are structurally conserved and very common in eukaryotes, the details of the loop regions and solvent-exposed strands are needed to provide functional information. The aim is to determine the structure of the mammalian ATG16L1 WD-40 domain, to confirm the organization of WD40-repeats and reveal the atomic and molecular details, providing a deeper insight into the role of this particular domain in the context of autophagy.

Another objective of this project is functional characterization of the WD-40 ATG16L1 domain. Mapping of WD-40 of ATG16L1 as a DCAF would confirm a direct interaction between DDB1 and ATG16L1. Moreover, ATG16L1 has been reported to interact with CARD1 domain of NOD2, CARD domain of NOD1, cytosolic domain of TLR2, and with the cytosolic domain of TMEM59. The aim is to analyze these interactions *in vitro*, and, if confirmed, co-crystallize interacting domains with the WD-40 of ATG16L1 to reveal accurate interaction surfaces. Only a crystal structure of the complex would reveal exactly which residues of the WD-40 propeller and its interaction partners are responsible for the interaction.

3 Materials and Methods

3.1 Materials

3.1.1 Enzymes

All enzymes, listed in the Table 2, were used according to the manufacturer's instructions.

Enzyme	Manufacturer
Calf Intestinal Phosphatase (CIP)	New England Biolabs
Restriction enzymes	New England Biolabs
T4 DNA ligase	New England Biolabs
KOD Polymerase	Novagen
Standards	
Smart Ladder (DNA)	Eurogentech
PageRuler unstained protein ladder	Fermentas
PageRuler Plus prestained protein ladder	Fermentas
2x Red PCR Master Mix	pjk

Table 2: Enzymes and standards used in this study, with their respective manufacturers

3.1.2 Kits

All kits, listed in Table 3, were used according to the manufacturer's instructions.

Kits	Manufacturer
QIAprep® Spin Miniprep Kit	Qiagen
QIAquick® PCR Purification Kit	Qiagen
QIAquick® Gel Extraction Kit	Qiagen
PureYieldTM Plasmid MidiPrep System	Promega
Proti-Ace	Hampton Reseach
Proti-Ace II	Hampton Reseach
Strep-tag® AP Detection Kit	Iba

Table 3: Kits used in this study, with their respective manufacturers DNA constructs

3.1.2.1 cDNA

Complementary DNA (cDNA) used as for amplifying genes of interests in this study are listed in Table 4.

cDNA	plasmid	clone ID	Supplier
Hs_ATG16L1	Pincy	I1382-8397581/ LIFESEQ2507216	Open Biosystems
Mm_ATG16L1	pYX-Asc	MMM1013-9498451	Open Biosystems
Hs_NOD2	pENTR223.1	HsCD00297035	DNASU
Hs_NOD1	pDONR223	CCSB 2450	CCSB
Hs_TMEM59	pENTR223	HsCD00509837	DNASU

Table 4: cDNA List of cDNA used as templates for cloning. Hs = Homo sapiens. Mm = Mus musculus

3.1.2.2 DNA constructs

All vectors in Table 5 have been designed and cloned previously, and used in this work.

Gene	expression system	Vector	Tag	resistance	reference
drDDB2	Insect	pFastBac-derived	N-His6	Amp	(Scrima *et al.* 2008)
HsDDB1_1-1104	Insect	pFastBac-derived	N-His6	Amp	(Scrima *et al.* 2008)
HsATG16L1_T300A_M1-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_T300A M1-Y607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L1_wt M1-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt M1-Y607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L1_wt K266-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt G284-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt G303-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_T300A K266-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_T300A G284-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt K266-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt G284-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_wt G303-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_T300A K266-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_T300A G284-Y607	Insect	pFastbacDual-derived	N-His6	Amp	
PSS013 HsNod2_M1-P209	insect/mammalian	pFlpBtMII (F13/F14)	N-Strep	Amp	
pSS001 HsNod2_M1-L1040	insect/mammalian	pFlpBtMII (F13/F14)	N-Strep	Amp	

Table 5: List of constructs used from previous work in our lab. Amp = ampicillin. Dr = Danio rerio. Hs = Homo sapiens

Gene	expression system	vector	Tag	resistance	Reference
hsLC3A	*E. coli*	pGEX4T	N-GST	Amp	Trond Lamark
hsLC3B	*E. coli*	pGEX4T	N-GST	Amp	Trond Lamark
hsGABARAP	*E. coli*	pGEX4T	N-GST	Amp	Trond Lamark
hsGABARAP L1	*E. coli*	pGEX4T	N-GST	Amp	Trond Lamark
hsGABARAP L2	*E. coli*	pGEX4T	N-GST	Amp	Trond Lamark

Table 6: List of constructs kindly provided by Trond Lamark from the Department of Medical Biology, The Artic University of Norway. Amp = ampicillin. Hs = Homo sapiens

Constructs listed in Table 7 are designed and cloned independently for the purpose of this work. First multiple cloning site (MCSI) has NotI/KpnI restriction sites (unless otherwise noted) used for incorporation of genes of interest.

Gene	expression system	Vector	tag	resistance	Reference
MmATG16L_wt_1-607	Insect	pFastbacDual-derived	N-His6	Amp	
MmATG16L_wt_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	
MmATG16L1_wt_P311-P607	Insect	pFastbacDual-derived	N-His6	Amp	
MmATG16L1_wt_P311-P607	Insect	pFastbacDual-derived	N-Strep	Amp	
MmATG16L1_wt_A320-P607	Insect	pFastbacDual-derived	N-His6	Amp	
MmATG16L1_wt_A320-P607	Insect	pFastbacDual-derived	N-Strep	Amp	
MmATG16L1_wt_P271-P607	Insect	pFastbacDual-derived	N-His6	Amp	
MmATG16L1_wt_P271-P607	Insect	pFastbacDual-derived	N-Strep	Amp	
MmATG16L1_wt_P231-P607	Insect	pFastbacDual-derived	N-His6	Amp	
MmATG16L1_wt_P231-P607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L_A300D_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	

Gene	expression system	Vector	tag	resistance	Reference
HsATG16L_A300E_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L_S287E_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L_S290E_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L_S304E_1-607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L_A320-607	Insect	pFastbacDual-derived	N-Strep	Amp	
HsATG16L1_266-607_SER1 (K469A/K470A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_266-607_SER2 (E354A/K355A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_266-607_SER3 (K306A/E307A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_266-607_SER4 (E569A/K570A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_284-607_SER1 (K469A/K470A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_284-607_SER2 (E354A/K355A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_284-607_SER3 (K306A/E307A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_284-607_SER4 (E569A/K570A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_303-607_SER1 (K469A/K470A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_303-Y607_SER2 (E354A/K355A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_303-P607_SER3 (K306A/E307A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_303-607_SER4 (E569A/K570A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsLC3C	*E. coli*	pGEX4T	N-GST	Amp	
HsATG16L1_wt_1-607	insect/mammalian	pFlpBtMII (wt/F3)	N-FLAG-HA	Amp	
HsATG16L1_303-607	insect/mammalian	pFlpBtMII (wt/F3)	N-FLAG-HA	Amp	
HsATG16L1_1-307	insect/mammalian	pFlpBtMII (wt/F3)	N-FLAG-HA	Amp	

Gene	expression system	Vector	tag	resistance	Reference
HsATG16L1_303-P607	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_303-Y607_SER2 (E354A/K355A)	Insect	pFastbacDual-derived	N-His6	Amp	
HsATG16L1_1-607	insect/mammalian	pFlpBtMII (F13/F4)	N-myc	Amp	
HsATG16l1_1-307	insect/mammalian	pFlpBtMII (F13/F4)	N-myc	Amp	
HsDDB1_M1-H1140	insect/mammalian	pFlpBtMII (wt/F3)	N-Strep	Amp	
HsNOD2_1-209	*E. coli*	pGEX-derived	N-GST	Amp	
HsNOD2_23-209	*E. coli*	pGEX-derived	N-GST	Amp	
HsNOD2_1-209	*E. coli*	pCOLA-Duet	N-Strep	Kan	
HsNOD2_23-209	*E. coli*	pCOLA-Duet	N-Strep	Kan	
HsNOD2_1-112	*E. coli*	pCOLA-Duet	N-Strep	Kan	
HsNOD2_1-121	*E. coli*	pCOLA-Duet	N-Strep	Kan	
HsNOD2_23-121	*E. coli*	pCOLA-Duet	N-Strep	Kan	
HsNOD2_28-119-Thrombin-218	*E. coli*	pGEX-derived	N-GST	Amp	
HsNOD2_28-218	*E. coli*	pGEX-derived	N-GST	Amp	
HsTMEM59_263-323	*E. coli*	pGEX-derived	N-GST	Amp	
HsTMEM59_263-323	*E. coli*	pET15b-derived	C-GST	Amp	
HsNOD1_1-106	*E. coli*	pET15b-derived	C-GST	Amp	
HsTMEM59_263-281	*E. coli*	pGEX-derived	N-GST	Amp	

Table 7: List of constructs cloned for this work. Amp = ampicillin; Kan = kanamycin. Hs = Homo sapiens. Mm = Mus musculus.

3.1.3 Oligonucleotides

Oligonucleotides used are listed in the Table 8.

oligonucleotide name	sequence (5’ -> 3’)	Organism	RE
HsATG16L_N_M1_f	aagaatgcggccgcatgtcgtcgggcctccgcgc	*Homo sapiens*	NotI
HsATG16L_Y607_K_r	ggggtacctcagtactgtgcccacagcac	*Homo sapiens*	KpnI
HsATG16L_Kdel_f	gaagtgagggttccagctactgc	*Homo sapiens*	none
HsATG16L_Kdel_r	gcagtagctggaaccctcacttc	*Homo sapiens*	none
HsATG16_K266_N_f	aagaatgcggccgcaagcgactctcgcagcctgc	*Homo sapiens*	NotI
HsATG16_G284_N_f	aagaatgcggccgcgggagacgctctgtctcttcc	*Homo sapiens*	NotI
HsATG16_G303_N_f	aagaatgcggccgcggttctggtaaagaagtgaggg	*Homo sapiens*	NotI
MmATG16L_P311-P607_f	aagaatgcggccgcccaactactgcctcgtatgtc	*Mus musculus*	NotI
MmATG16L_A320-P607_f	aagaatgcggccgcgcgcatgacggagaggtcaac	*Mus musculus*	NotI
MmATG16L_P271-P607_f	aagaatgcggccgccctgctggaggccttctgg	*Mus musculus*	NotI
MmATG16L_P231-P607_f	aagaatgcggccgccctctacctgttgaacagg	*Mus musculus*	NotI
MmATG16L_int_seq	gaatgaagcaaagatttcgg	*Mus musculus*	NotI
HsDDB1_G490_f	gtcattgatgtcaagttcctatatgg	*Homo sapiens*	none
HsDDB1_G961_f	gttgtgtttgtcgggtctcgcc	*Homo sapiens*	none
HsDDB1_G2221_f	gaagtccaagacacgagtg	*Homo sapiens*	none
HsDDB1_G2901_f	ggctgaaaatgcctttaacttgtttg	*Homo sapiens*	none
HsDDB1_T1601_f	tggaacatgaagtggcttgcttg	*Homo sapiens*	none
HsATG16L_A300D_f	gtggatgatcatcctggttctggtaaag	*Homo sapiens*	none
HsATG16L_A300D_r	aggatgatcatccacattgtcctggggg	*Homo sapiens*	none
HsATG16L_A300E_f	gtggatgaacatcctggttctggtaaag	*Homo sapiens*	none
HsATG16L_A300E_r	aggatgttcatccacattgtcctggggg	*Homo sapiens*	none
DDB1_F598_r	cttttctcggagagacacctcatagg	*Homo sapiens*	none
HsATG16L_S287E_f	cgcgaagtctcttccttcccagtcccccaggac	*Homo sapiens*	none
HsATG16L1_S287E_r	gaaggaagagacttcgcgtctcccaaagatattag	*Homo sapiens*	none
HsATG16L1_S290E_f	tctgtctctgaattcccagtcccccaggac	*Homo sapiens*	none
HsATG16L_S290E_r	gaattcagagacagagcgtctcccaaagatattag	*Homo sapiens*	none
HsATG16L_S304E_r	ctttaccttcaccaggatgagcatccacattg	*Homo sapiens*	none
HsATG16_2xE_f	cgcgaagtctctgaattcccagtcccccaggac	*Homo sapiens*	none

oligonucleotide name	**sequence (5' -> 3')**	**Organism**	**RE**
HsATG16_2xE_r	gaattcagagacttcgcgtctcccaaagatattag	*Homo sapiens*	none
HsATG16_3xE_f	cgcgaagtcgaagaattcccagtcccccaggac	*Homo sapiens*	none
HsATG16_3xE_r	gaattcttcgacttcgcgtctcccaaagatattag	*Homo sapiens*	none
HsATG16_A320_f	aagaatgcggccgcgcacatgatggggaagtcaac	*Homo sapiens*	none
HsATG16L_S289E_f	tctgtcgaatccttcccagtcccccaggacaatg	*Homo sapiens*	none
HsATG16L_S289E_r	gaaggattcgacagagcgtctcccaaagatattagtg	*Homo sapiens*	none
HsATG16L_S304E_f	cctggtgaaggtaaagaagtgagagtac	*Homo sapiens*	none
HsATG16L1_WD40_SER1_f	gacgccgccattcgtttctgggacattcgatcag	*Homo sapiens*	none
HsATG16L1_WD40_SER1_r	cgaatggcggcgtcaaaatgtccactcattacacattg	*Homo sapiens*	none
HsATG16L1_WD40_SER2_f	ggagccgcctgtgagttcaagggttccctatc	*Homo sapiens*	none
HsATG16L1_WD40_SER2_r	acaggcggctccaaatacttcccaaagcttaacc	*Homo sapiens*	none
HsATG16L1_WD40_SER3_f	gtggccgccgttctttcaaagcagcacagctcatc	*Homo sapiens*	none
HsATG16L1_WD40_SER3_r	gaacggcggccactttccctgtgagcacactc	*Homo sapiens*	none
HsATG16L1_WD40_SER4_f	ggtgccgccgtgagggttccagctactgcc	*Homo sapiens*	none
HsATG16L1_WD40_SER4_r	cacggcggcaccagaaccaggatgggtatcc	*Homo sapiens*	none
HsNod1_N_M1_f	aagaatgcggccgcatggaagagcagggccacagtg	*Homo sapiens*	NotI
HsNOD1_K_106E_r	ggggtaccctccagcagccaaggcctgag	*Homo sapiens*	KpnI
HsTMEM59_N_T263_f	aagaatgcggccgcacagctgtggagcagtatgttcc	*Homo sapiens*	NotI
HsTMEM59_K_I323_stop_r	ggggtaccttaaatttcagaatgagcaagattcac	*Homo sapiens*	KpnI
HsTMEM59_N_T263_start_f	aagaatgcggccgcatgacagctgtggagcagtatgttcc	*Homo sapiens*	NotI
HsTMEM59_K_I323_r	ggggtaccaatttcagaatgagcaagattcacttttg	*Homo sapiens*	KpnI
HsLC3C_B_f	cgggatccatgccgcctccacagaaaatcc	*Homo sapiens*	BamHI
HsLC3C_N_r	ttcttagcggccgctcgagttagagaggattgcagggtctgtcctcaaggc	*Homo sapiens*	NotI

*Table 8: List of oligonucleotides used for gene amplification. (RE = **re**striction sites)*

3.1.4 Sequencing primers

Primers used for sequencing are listed in Table 9.

primer name	sequence (5’ -> 3’)
pFB1_5’_f	tattccggattattcatacc
pFB1_3’_r	cctctacaaatgtggtatgg
T7_f	taatacgactcactataggg
T7_r	gctagttattgctcagcgg
pGEX_f	atagcatggcctttgcagg
pGEX_r	gagctgcatgtgtcagagg
DuetUP1	ggatctcgacgctctccct
DuetDOWN1	gattatgcggccgtgtacaa
seq_pFliptomam_r	cctcaagacccgtttagaggcc
pFlpBtM_seq_for_II	cattttatttacaatcaggatccg

Table 9: List of primers used for in-house sequencing of genes used in this study

3.1.5 Bacterial strains and cell lines

Bacterial strains are listed in Table 10. DH5α and Rosetta 2(DE3) are used for *E. coli* expression system. DH5α are used for cloning while Rosetta 2(DE3) cells are used for protein expression and production. DH10EMBacY cells are used to generate baculovirus for insect cell protein expression.

strain	**Genotype**	**manufacturer**
DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–	Invitrogen
Rosetta 2(DE3)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)	Novagen
DH10EMBacY	F-mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80(lacZ) ΔM15ΔlacX74 recA1 araD139 Δ(ara,leu)7697 galU galK l- rpsL nupG /BacloxP/pBADZ-His6Cre/pMON7124	EMBL

Table 10: List of bacterial strains

3.1.6 Insect cell lines

IPLB-SF-21AE (Sf-21) isolated from ovaries of an armyworm *Spodoptera frugiperda* (Vaughn *et al.* 1977) and adapted for growth in suspension. Sf-21 is used for generation of baculovirus and its propagation.

BTI-Tn-5B1-4 (High Five) was originally isolated from ovaries of a cabbage looper *Trichoplusia ni* (Wickham and Nemerow 1993) and is commercially available as High Five™ from Invitrogen. This cell line is used for protein production.

3.1.7 Mammalian cell lines

HEK293-6E is a subclone of a human embryonic kidney epithelial cell line HEK293 (Graham, Smiley, Russell, & Nairn, 1977). This subclone is adapted to suspension and carries a truncated Epstein-Barr nuclear antigen 1 (EBNA1) (Durocher *et al.* 2002). This cell line is used for transient test expression.

3.1.8 Antibiotics

Antibiotics listed in Table 11 were used to select for plasmids caring the corresponding resistance gene. In addition to antibiotic section, Blue-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isopropyl β-D-1-thiogalactopyranoside) were used for blue-while screening of DH10EMBacY cells.

Antibiotics and chemicals	**Final concentration (µg/ml)**
Ampicillin (Amp)	100
Chloramphenicol (Cm)	25
Gentamicin (Gm)	7
Kanamycin (Kan)	50
Tetracyclin (Tet)	10
Blue-Gal	100
IPTG	40

Table 11: List of antibiotics and chemicals used for growth and selection of desired colonies

3.1.9 Antibodies

Antibodies used for detection of proteins and tags are listed in the Table 12.

Antibody	**Manufacturer**
ATG5	GeneTex
GST-HRP conjugate	AmershamBioscience
LC3A/B	Cell Signaling
ATG16L	Sigma
ImmunoPure Avidin, AP conjugated	ThermoScientific

Table 12: List of antibodies and their manufacturers

3.1.10 Columns and resins used in this study

A list of affinity resins and ion exchange and gel filtration columns used in this study is presented in Table 13.

Resin	Manufacturer
Ni Sepharose® 6 Fast Flow	GE Healthcare
Glutathione Sepharose® 4 Fast Flow	GE Healthcare
Strep-Tactin® Sepharose® 50% suspension	Iba
Strep-Tactin® Macroprep® 50% suspension	Iba
Strep-Tactin® Superflow® high capacity 50% suspension	Iba
Pierce® Monomeric Avidin Agarose	Thermo Fisher Scientific
Anti-DYKDDDDK (FLAG) Tag (L5) Affinity Gel	Biolegend
Columns	**Manufacturer**
Source15Q ion exchange (8ml)	GE Healthcare
HiLoad 16/600 Superdex 200 prep grade	GE Healthcare
HiLoad 16/600 Superdex 200 prep grade	GE Healthcare
Superdex 200 10/30	GE Healthcare

Table 13: List of resins and columns used for protein studies in this work

3.1.11 Media and buffers

3.1.11.1 Media for eukaryotic systems

Media used for propagation of insect and mammalian cells are listed in Table 14.

Cell culture media	Cell lines	Manufacturer
Ex-Cell™ 405	HighFive	SAFC
Ex-Cell™ 420	Sf-21	SAFC
F17	HEK293-6E	Gibco

Table 14: List of media for insect and mammalian cells

F17 medium from HEK293-6E cells is supplemented by 0.025mg/L G418, 1g/L Pluronic F68 and 7.5mM L-Glutamine prior to use.

3.1.11.2 Media for prokaryotic systems

Bacterial culture media	Composition
Luria Bertani (LB)	10g/L tryptone, 5g/L, NaCl, 5g/L yeast extract
Terrific Broth (TB)	12g/L tryptone, 24g/L yeast extract, 4ml/L glycerol, buffered with 10% 0.17 M KH_2PO_4, 0.72 M K_2HPO_4
SOC-Medium	20g/L tryptone, 5g/L yeast extract, 0.5g/L NaCl, 2.5mM KCl, 10mM $MgCl_2$, 20mM glucose
TFB1	30mM $KOOCH_3$, 100mM RbCl, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% glycerol, buffered with 5M acetic acid to pH 5.8
TFB2	10mM MOPS, 75mM $CaCl_2$, 10mM $RbCl_2$, 15% glycerol, buffered with 50% (w/v) NaOH to pH 6.5

3.1.11.3 Buffers

Buffers	Composition
PBS buffer	137mM NaCl, 3mM KCl, 12mM Na_2HPO_4, 2mM K_2HPO_4, pH 7.4
PBS-T buffer	PBS buffer, 0.1% (v/v) Tween-20
AF4 running buffer	PBS, 0.003M sodium azide
Blocking buffer	5% (w/v) skim milk in PBS-T
2xDTT loading buffer	100mM Tris/HCl pH 6.8, 4% SDS, 20% glycerol, 200mM DTT, 0.2% (w/v) bromphenol blue
SDS-PAGE lower buffer (4x)	1.5M Tris/HCl, pH 8.8
SDS-PAGE running buffer	25mM Tris/HCl, 192mM glycine, 0.1% (w/v) SDS
SDS-PAGE sample buffer (2x)	1ml 1M Tris/HCl pH 6.8, 2.4mM glycerol, 0.8g SDS, 2mg Coomassie blue G-250, 0.31g DTT, add H_2O to 10ml
SDS-PAGE separating gel	10ml Polyacrylamide solution, 5ml 4x Lower buffer, 0.2ml SDS (10%), 4.7ml H_2O, 20 µl TEMED, 50µl APS (25%)
SDS-PAGE stacking gel	1.5ml Polyacrylamide solution, 2.5ml 4 x Upper buffer, 5.9ml H_2O, 15µl TEMED, 25µl APS (25%)

Buffers	Composition
SDS-PAGE staining solution	30% (v/v) ethanol, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie R-250
SDS-PAGE Upper buffer (4x)	0.5M Tris/HCl pH 6.8, 0.4% (w/v) SDS
TAE buffer	40mM Tris/HCl pH 8.2, 20mM sodium acetate, 1mM EDTA
DNA loading buffer (10x)	10mM Tris/HCl, pH 7.5, 0.05% (w/v) bromphenol blue, 1mM EDTA, 50% glycerol
Transfer buffer	20mM Tris pH 8.0, 192mM glycine, 15% methanol
CD buffer	10mM NaH_2PO_4, pH 8, 10mM NaCl, 5mM DTT, 1mM $MgCl_2$
Hank's Buffered Salt Solution (HBSS)	0.127M NaCl, 5.4mM KCl, 0.25mM Na_2HPO_4, 0.1g/L glucose, 0.44mM KH_2PO_4, 1.3mM $CaCl_2$, 1.0mM $MgSO_4$, 4.2mM $NaHCO_3$

3.1.12 Crystallization Screens

Crystallization screens used for initial screening and optimization are listed in Table 15.

Screen	Manufacturer
JCSG Core I-IV	Qiagen
AmSO4	Qiagen
Morpheus	Molecular Dimensions
Midas	Molecular Dimensions
Index	Hampton Research
Additive Screen	Hampton Research
Silver Bullet Screen	Hampton Research

Table 15: List of commercial crystallization screens used in this study, with their respective manufacturers

3.2 Methods

3.2.1 Molecular Cloning

3.2.1.1 Polymerase chain reaction (PCR)

Gene amplification was done by PCR using standard procedure. PCR reaction mix is represented in Table 16.

Component	Volume
DNA template (10 ng/µl)	1 µl
10xBuffer 1 for KOD Hot Start	5 µl
$MgSO_4$ (25 mM)	3 µl
dNTPs (2 mM each)	5 µl
Primer frw (10 pmol/µl)	2 µl
Primer rev (10 pmol/µl)	2 µl
KOD HotStart Polymerase	1 µl
MQ H_2O	31 µl

Table 16: Composition of PCR mix, final volume of 50 µl

The PCR-cycling parameters shown in Table 17 are used. Annealing temperature is chosen according to the melting temperature of primer pairs. Elongation time is adjusted depending on the length of the PCR product according to the manufacturer's instructions.

Step	T(°C)	Duration (s)	Repeat
Initial denaturation	95	120	
Denaturation	95	30	40
Annealing	55	15	
Elongation	72		
Final extension	72	420	
Hold	4	∞	

Table 17: PCR cycle stages and their duration

3.2.1.2 DNA agarose gel electrophoresis and gel extraction purification

PCR products are visualized on 0.8 % agarose gel (unless otherwise noted) with 0.5µg/ml ethidium bromide. Samples were mixed with 10xDNA loading dye. 1xTBE was used as running buffer. Agarose gels are visualized under UV light (Ultrospec 3000, 245nm). Extraction of PCR products from the agarose gel is done using the Qiagen Gel Extraction Kit according to the manufacturer‘s protocol.

3.2.1.3 Restriction digest

Cloning vectors and PRC products were digested at 37°C for 1hr with respective enzymes (Table 18) in 40µl reactions.

	Vector	**PCR product**
DNA amount	2 µl	x µl
NEB 10x Buffer 4	4 µl	4 µl
KpnI	1 µl	1 µl
NotI	1 µl	1 µl
MQ H_2O	32 µl	ad 40 µl

Table 18: Restriction reaction composition

Upon digestion, cloning vectors are additionally treated with CIP for 1hr, 37°C to dephosphorylate the ends and reduce the re-ligation rate.

Following restriction digest, PCR products are purified using QIAGEN PCR Purification Kit. Vectors are loaded on 0.8% agarose gel with 0.5µg/ml ethidium bromide and bands corresponding to the digested plasmid are cut out and purified using QIAGEN Gel Extraction Kit according to the manufacturer's protocol. Concentration of purified DNA is determined by NanoDrop spectrometer (ND-1000, PeqLab Biotechnology).

3.2.1.4 Ligations

Ligations are performed using 50-100ng of vector with 5 fold molar excess of insert. A final reaction volume of 20µl is set up with 2µl 10x T4 ligase buffer and 1µl T4 DNA ligase. As control, same reaction is set up with vector only, to assess the success of digestion. Ligation reactions are incubated at room temperature for 1hr.

3.2.1.5 Chemocompetent cells

An overnight 5ml culture of chemocompetent cells in inoculated into 250ml fresh LB media and grown at 37°C to OD_{600} of 0.5. Culture is chilled on ice for 10 min prior to centrifugation at 4°C, 300 rpm, 5 min. Bacterial pellet is carefully and thoroughly resuspended in 100 ml TFB1, chilled on ice for 10min and centrifuged at 4°C, 300rpm, 5min. This time pellet is resuspended in 10ml TFB2. Cells are shock frozen in liquid nitrogen in 100µl aliquots and stored at -80°C.

3.2.1.6 Transformation

Transformations are performed by heat shock. Chemocompetent *E. coli* (100µl) are incubated either with 10µl of the ligation reaction, or with 1µl of purified DNA for 30min on ice. Cells are heat shocked at 42°C for 1.5min, followed by 2min incubation on ice. Pre-warmed 800µl of SOC media is added to the cells and incubated at 37°C, 550rpm (Thermomixer 5436, Eppendorf). In case of DH5α and Rosetta 2(DE3), cells are incubated for 1hr; DH10EMBacY are incubated for 4hr. Bacteria are centrifuged at the table top centrifuge (Fresco21, ThermoScientific) for 1min, 14000rpm. Excess media is decanted and cells are plated on LB-Agar plates containing appropriate antibiotics.

3.2.1.7 Colony PCR

To confirm the presence of the insert, colony PCR was performed using sequencing primers. PCR reaction mix is represented in Table 19.

Component (per colony)	Volume
Colony	
2x PCR Red dye Mix	10 µl
Primer frw (10 pmol/µl)	1 µl
Primer rev (10 pmol/µl)	1 µl
MQ H_2O	8 µl

Table 19: PCR mix composition for colony PCR. Total reaction volume is 20 µl

The PCR-cycling parameters shown in Table 20 are used. Annealing temperature is chosen according to the melting temperature of primer pairs. Elongation time is adjusted depending on the length of the PCR product according to the manufacturer’s instructions.

Step	T(°C)	Duration (s)	Repeat
Initial denaturation	94	300	
Denaturation	95	30	26
Annealing	55	30	
Elongation	72		
Final extension	72	420	
Hold	4	∞	

Table 20: Colony PCR cycle stages and their duration

Colonies positive for inserts are inoculated in 5ml LB media supplied with appropriate antibiotics and grown overnight at 37°C, 130rpm in shaking incubator (Infors Af 110).

3.2.1.8 Plasmid purification

Plasmid purification is done using QIAprep Spin MiniPrep Kit for DNA extraction from overnight cultures according to the manufacturer’s protocol. Purified plasmids are sent for in-house sequencing. Plasmids are kept at -20°C for long term storage, and glycerol stocks are kept at -80°C.

For transient transfection of HEK293-6E cells, plasmids are prepared according to the PureYield^{TM} Plasmid MidiPrep System.

3.2.1.9 Isolation of bacmid

The bacmid DNA is prepared according to the protocol in the BAC-TO-BAC® system manual from Life Technologies. 5ml LB media supplemented with 50µg/ml kanamycin, 7µg/ml gentamicin, 10µg/ml tetracycline are inoculated with a single white, isolated colony and incubated overnight at 37°C, 130 rpm in a shaking incubator (Infors Af 110). Following 3 min centrifugation at 14000 rpm, pellet from 2ml of culture is suspended in 300µl P1 solution (QIAgen). After addition of 300µl P2 solution (QIAgen) and gentle agitation, the mixture is incubated at room temperature for 5 min. Another 300µl of P3 solution (QIAgen) is added, mixed and incubated on ice for 5-10 min. Solution is cleared by centrifugation at 14000 rpm, 10 min, 4°C. The supernatant is transferred into a tube containing 800 µl isopropanol, mixed gently and incubated 5-10 min on ice. Following another centrifugation at 14000 rpm, 10 min, 4°C and supernatant is decanted and the pellet is washed with 500µl 70% ethanol. Ethanol is removed by centrifugation at 14000 rpm, 5 min, 4°C followed by air drying at room temperature. Finally, DNA pellet is dissolved in 40µl EB-buffer (QIAgen) and stored at -20°C.

3.2.2 Cell culture methods

3.2.2.1 Transfection of Sf-21 with bacmid DNA and virus amplification

Sf-21 and HighFive cells are cultivated at 27°C, 100 rpm and passaged every two to three days to maintain cell density between $3x10^5$ and $3x10^6$ cells/ml. To generate recombinant viruses, 1ml of 0.6-$0.8x10^6$ adherent Sf-21 cells is seeded on 6-well plates and transfected with recombinant bacmids. For each transfection reaction, 10µl Superfect (QIAGEN) and 5-10µl purified bacmid are mixed with 150µl ExCell 420 media and incubated at room temperature for 10 min. Following the incubation, total volume is brought to 1ml with 5% FCS/ExCell420 mixture and gently added to the cells. Cells are incubated at 27°C. The supernatant containing first generation viruses (P0) is harvested six days post transfection.

For each subsequent generation (P1-P3), 10% of previous virus stock is added to 0.6-$0.8x10^6$ cells/ml in serum free suspension culture and incubated at 27°C, 100 rpm. Infection is determined by measuring yellow fluorescence on GUAVA EasyCyte Mini (Merck Milipore) and increase of cell size by CASY. Virus is harvested typically 3 days post infection, by centrifugation at 2000 rpm, 4 min. Supernatant is stored at 4°C with 5% FCS.

3.2.2.2 Cell size determination

Increase of cell diameter is one of the key criteria of assessment of insect cell infection. Cell diameter can be measured by CASY Cell Counter (Innovatis). The measurements were performed according to the manufacturer's guidelines.

3.2.2.3 Preparation of baculovirus infected insect cells (BIIC)

The baculovirus infected insect cells (BIIC) method is developed as an improvement for long term storage of recombinant baculoviruses (Wasilko *et al.* 2009). BIIC stocks are made by infecting 300 ml Sf-21 cells ($1x10^6$ cells/ml) in serum free media with 20% of P2 or P3 virus. The culture is incubated at 27°C, 100 rpm, and cell diameter and vitality are monitored daily. Cells are harvested latest 48 hr post infection, once the cell diameter increased by 3-5µm (10-20% of the size on the day of infection) and the vitality was >85%. Harvesting is done by centrifuging the culture at 180 rpm for 4 min. Supernatant is decanted and the pellet is resuspended in freshly prepared freezing media (90% ExCell 420, 10% DMSO) at $1x10^7$ cells/ml. Each cryo-tube (Nunc) is filled with 1.5ml suspension. The cryo-vials are stored in Mr. Frosty (Nalgene) at -80°C for 24 hr to provide a constant cooling rate of 1°C/min. Subsequently, cryo-vials are stored in the vapor phase over liquid nitrogen for long term storage.

3.2.2.4 Transient transfection of HEK293-6E cells with plasmid DNA

HEK293-6E cells are cultivated in supplemented F17 media at 37°C, 100 rpm, 5% CO_2 by Nadine Konisch. For transfection 2µg of plasmid DNA per $1x10^6$ cells were incubated in 200µl F17 media according to the standard transfection conditions as described by Bollin *et al.* (Bollin *et al.* 2011) The HEK293-6E cells were seeded in a total volume of 1.8ml at a concentration of $0.6x10^6$ cells/ml and afterwards the 200µl transfection mix was added.

3.2.3 Protein expression and purification

3.2.3.1 Protein test expression from *E. coli*

Plasmids are transformed into Rosetta 2 (DE3) cells and protein expression is tested in TB media (unless otherwise specified) at 20°C and 37°C. Overnight cultures are used to grow a larger culture to OD_{600}=0.5-0.6. Protein expression is induced by 50 or 500µM IPTG and bacterial pellets are collected after 3 hr or overnight to test protein expression under corresponding conditions. Cells are harvested (10ml/condition), resuspended in 1ml lysis buffer (50mM Tris pH 8, 200mM NaCl, 1mM PMSF, 5mM BME) and lysed by sonication. Following 45 min centrifugation at 14000 rpm, 4°C, clear lysate is incubated with 30µl appropriate beads and incubated for 1 hr, 4°C, on a rocking platform. Beads are subsequently washed and dissolved in 2x DTT loading buffer for SDS-PAGE analysis. Additionally, samples are taken from pellet and supernatant prior to addition of beads.

3.2.3.2 Protein test expression from insect cells

Plasmids are transformed into EMBacY cells and used for virus production and amplification in Sf-21 cells. After passage 2 (P2), Sf-21 cells are tested for protein expression. Cells are harvested (10ml), resuspended in 1ml lysis buffer (50mM Tris pH 8, 200mM NaCl, 1mM PMSF, 5mM BME, 0.1% Triton X-100) and lysed by sonication. Following 45 min centrifugation at 14000 rpm, 4°C, clear lysate is incubated with 30 µl appropriate beads and incubated for 1hr, 4°C, on a rocking platform. Beads are subsequently washed and dissolved in 2x DTT loading buffer for SDS-PAGE analysis.

3.2.3.3 Protein production and purification of ATG16L1

3.2.3.3.1 Protein production and purification of ATG16L1 full length

Human ATG16L1 was cloned by PCR using cDNA from Open Biosystems (Table 4) as a template. Petra Völler manipulated this clone to eliminate internal KpnI site by introducing a silent mutation. ATG16L1 was cloned into *E. coli* and insect cell expression vectors by Petra Völler and tested for expression in *E. coli*. I transformed EMBacY cells, prepared bacmid and transfected Sf-21 insect cells for virus production. After P2 test expression and subsequent production of P3 virus, HighFive cells are used for protein production. Recombinant virus harboring full length ATG16L1 was not stable and thus BIIC stock is generated and used for infection of HighFive cells.

Strep-ATG16L1 full length:

HighFive cells (0.5-0.8x10^6 cells/ml) (unless otherwise noted) are infected with 5-10% P3 virus or with one BIIC aliquot/1L culture and incubated at 27°C, 100 rpm for 72 hr. Cells are harvested 3 days post infection in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 2000 rpm, 5 min, 4°C. Pellet is resuspended in Lysis Buffer (Table 21). Cells are lysed by Bandelin SonoPlus HD200 sonicator and pelleted by centrifugation at 60000 rpm, 1 hr, 4°C in 70 Ti rotor (Beckman Coulter).

The supernatant is incubated with 2-3ml of *Strep*-Tactin® SuperFlow® High Capacity resin (IBA) (unless otherwise noted) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins are washed by 10 column volumes (CV) of Washing Buffer (Table 21). The target protein is eluted with 5x 5ml Elution Buffer (Table 21).

Affinity Chromatography

Buffers	**Composition**
Lysis Buffer	50mM Tris pH 8, 200 mM NaCl, 0.1% Triton X-100, 5mM BME and Protease Inhibitor Cocktail tablet (Sigma)
Wash Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME
Elution Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME, 5mM desthiobiotin

Table 21: List of buffers used for purification of Strep-tagged ATG16L1 full length

FLAG-ATG16L1 full length:

HighFive cells (0.5-0.8x10^6 cells/ml) are infected with 5-10% P3 virus and incubated at 27°C, 100 rpm for 72 hr. Cells are harvested 3 days post infection in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 2000 rpm, 5 min. Pellet is resuspended in Lysis Buffer (Table 21). Cells are lysed by Bandelin SonoPlus HD200 sonicator and pelleted by centrifugation at 60000 rpm, 1 hr, 4°C in 70 Ti rotor (Beckman Coulter).

The supernatant is incubated with 4ml of anti-DYKDDDDK tag (L5) affinity gel (BioLegend) and purified according to the manufacturer's technical bulletin.

3.2.3.3.2 Protein production and purification of His-tagged ATG16L1 WD-40 domain constructs

Construct boundaries for WD-40 propeller domain of ATG16L1 were based on secondary structure predication of human ATG16L1 gene by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) and designed by Dr. Andrea Scrima. WD-40 constructs were cloned into *E. coli* and insect cell expression vectors by Petra Völler. Petra Völler tested expression in *E. coli*. I transformed the insect cell constructs into EMBaCY cells, prepared bacmid and transfected Sf-21 cells for virus production. After P2 test expression and subsequent production of P3 virus, HighFive cells are used for protein production.

HighFive cells (0.5-0.8x10^6 cells/ml) are infected with 5-10 % P3 virus and incubated at 27°C, 100 rpm for 72 hr. Cells are harvested 3 days post infection in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 2000 rpm, 5 min. Pellet is resuspended in Lysis Buffer (Table 2.18). Cells are lysed by Bandelin SonoPlus HD200 sonicator and pelleted by centrifugation at 60000 rpm, 1 hr, 4°C in 70 Ti rotor (Beckman Coulter). The supernatant is incubated with 2-3 ml of Ni Sepharose® 6 Fast Flow (GE Healthcare) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins were washed by 25 CV of Washing Buffer (Table 22). The target protein was eluted with 20 ml Elution Buffer (Table 22).

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	50mM Tris pH 8, 200mM NaCl, 0.1% Triton X-100, 5mM BME and Protease Inhibitor Cocktail tablet (Sigma)
Wash Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME, 10mM Imidazole
Elution Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME, 150mM Imidazole
Size Exclusion Chromatography	
Running Buffer	20mM HEPES pH7.4, 200mM NaCl, 5mM DTT

Table 22: List of buffers used for purification of his-tagged ATG16L1 WD-40 domain constructs

Eluted protein is concentrated (Vivaspin®, 30000 MWCO, GE Healthcare) to ~1ml and applied on equilibrated Superdex75 16/60 (GE Healthcare) column connected to Äkta Purifier (GE Healthcare). All fractions of affinity and size exclusion chromatography are analyzed on 12% SDS-PAGE. Fractions containing pure protein are concentrated to ~20mg/ml, aliquoted, shock-frozen in liquid nitrogen and stored at -80°C.

3.2.3.3.3 Protein production and purification of DDB1

Human DNA damage-binding protein 1 is cloned and purified as presented in Scrima *et al.* 2008. His-tagged DDB1 is expressed from insect cells using standard protocol. Shortly, HighFive cells (0.5-0.8x10^6 cells/ml) are infected with 5-10% P3 virus and incubated at 27°C, 100 rpm for 72 hr. Cells are harvested 3 days post infection in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 2000 rpm, 5 min. Pellet is resuspended in Lysis Buffer (Table 23). Cells are lysed by Bandelin SonoPlus HD200 sonicator and pelleted by centrifugation at 60000 rpm, 1 hr, 4°C in 70 Ti rotor (Beckman Coulter). The supernatant is incubated with 2-3ml of Ni Sepharose® 6 Fast Flow (GE Healthcare) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins were washed by 25 CV of Washing Buffer (Table 23). The target protein was eluted with 20 ml Elution Buffer (Table 23).

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	50mM Tris pH 8, 200mM NaCl, 0.1% Triton X-100, 5mM BME and Protease Inhibitor Cocktail tablet (Sigma)
Wash Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME, 10mM Imidazole
Elution Buffer	50mM Tris pH 8, 200mM NaCl, 5mM BME, 100mM Imidazole

Ion Exchange Chromatography	
Low Salt Buffer	50 mM Tris pH 8, 5mM DTT
High Salt Buffer	50 mM Tris pH 8, 1 M NaCl, 5 mM DTT
Size Exclusion Chromatography	
Running Buffer	50 mM HEPES pH 7.4, 200 mM NaCl, 5 mM DTT

Table 23: List of buffers used for purification of His-tagged human DDB1

Elution fractions containing target protein are diluted with Low Salt Buffer (Table 23) to low salt (50mM NaCl) and loaded on Source15Q ion exchange column. After a washing step with 50 CV with Low Salt Buffer, protein is eluted with a 200-300mM NaCl gradient over six CV. Fractions of ion exchange are tested on 12% SDS PAGE gels and those containing target protein are concentrated (Vivaspin®, 100000 MWCO, GE Healthcare) to 1ml and loaded on equilibrated Superdex 200 16/60 connected to Äkta Purifier (GE Healthcare) as a final purification step. Purified DDB1 is concentrated to 3mg/ml, aliquoted, shock frozen in liquid nitrogen and stored at -80°C for interaction studies.

3.2.3.3.4 Protein production and purification of NOD2 CARD domains

Caspase recruitment domains (CARDs) of NOD2 are cloned in Step-tagged pCOLADuet and GST-tagged pGEX-derived vector and tested in *E. coli* for optimal expression condition. Strep-tagged protein showed highest yield when expressed in TB buffer at 20°C, 500µM IPTG for 3 hr. After 3 hr, cells are harvested in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 5000 rpm, 10 min. Pellet is resuspended in Lysis Buffer (Table 24) and the suspension was passed through a homogenizer (Avestin EmulsiFlex-C3) twice. The lysate was cleared by centrifugation at 16000 rpm, 45 min, 4°C.

The supernatant is incubated with 5ml of *Strep*-Tactin® Macroprep resin equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins are washed by 10 CV of Washing Buffer (Table 24). The target protein was eluted with 6x 5ml Elution Buffer (Table 24).

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME, Protease Inhibitor Cocktail tablet (Sigma)
Wash Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME
Elution Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME, 5mM desthiobiotin
Size Exclusion Chromatography	
Running Buffer	10mM HEPES pH 7.4, 100mM NaCl, 5mM DTT

Table 24: List of buffers used for purification of Strep-tagged NOD2 CARDs

Eluted protein is concentrated (VivaspinTM, 3000 MWCO, GE Healthcare) to ~1ml and applied on equilibrated Superdex75 16/60 (GE Healthcare) column connected to Äkta Purifier (GE Healthcare). All fractions of affinity and size exclusion chromatography were analyzed on 15% SDS-PAGE. Fractions containing pure protein are concentrated, aliquoted, and shock-frozen in liquid nitrogen and stored at -80°C.

3.2.3.3.5 Protein production and purification of TMEM59 cytosolic domain

TMEM59 T263-E281 peptide is cloned into pGEX vector. In addition, full cytosolic domain (T263-323) was cloned into an N-terminal GST pGEX vector, as well as in a C-terminal GST pET15b derived vector. All constructs were tested in *E. coli* for optimal conditions for protein expression and the highest yield for all of them was observed when expressed in TB media for 3 hr, 37°C, 500 µM IPTG. After 3 hr of expression, cells from 4 L TB culture are harvested in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 5000 rpm, 10 min. Pellet is resuspended in Lysis Buffer (Table 25) and the suspension was sonicated for 30 min, 1 sec pulse, 4 sec pause. The lysate was cleared by centrifugation at 16000 rpm, 45 min, 4°C.

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	1xPBS, 5mM BME, Protease Inhibitor Cocktail tablet (Sigma)
Wash Buffer	1x PBS, 5mM BME
Elution Buffer	1x PBS, 5mM BME, 20mM L-glutathione reduced
Size Exclusion Chromatography	
Running Buffer	1x PBS, 5mM DTT

Table 25: List of buffers used for purification of GST-TMEM59

The supernatant is incubated with 5ml of Glutathione Sepharose® 4 Fast Flow (GE Healthcare) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins are washed by 10 CV of Washing Buffer (Table 25). The target protein is eluted with 20 ml Elution Buffer (Table 25). Eluted protein is concentrated (VivaspinTM, 3000 MWCO, GE Healthcare) to ~1ml and applied on equilibrated Superdex75 16/60 (GE Healthcare) column connected to Äkta Purifier (GE Healthcare). All fractions of affinity and size exclusion chromatography are analyzed on 15% SDS-PAGE. Fractions containing pure protein are concentrated, aliquoted, shock-frozen in liquid nitrogen and stored at -80°C.

3.2.3.3.6 Protein production and purification of TLR2 TIR domain

Toll/interleukin receptor (TIR) domain of TLR2 is cloned in pET15b derived vector with His tag. His-tagged TIR TLR2 protein resulted highest yield when expressed at 37°C, 3 hr, 50 µM IPTG in LB buffer. After 3 hr of expression, cells are harvested in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 5000 rpm, 10 min. Pellet is resuspended in Lysis Buffer (Table 26) and the suspension is passed through a homogenizer (Avestin EmulsiFlex-C3) twice. The lysate is cleared by centrifugation at 16000 rpm, 45 min, 4°C.

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	20mM MOPS pH 7, 200mM NaCl, 5% Glycerol, 5mM BME, Protease Inhibitor Cocktail tablet (Sigma), DNase, lysozyme, 0.5mM $MgSO_4$
Wash Buffer	20mM MOPS pH 7, 200mM NaCl, 5% Glycerol, 5mM BME
Elution Buffer	20mM MOPS pH 7, 200mM NaCl, 5% Glycerol, 150mM Imidazole, 5mM BME
Size Exclusion Chromatography	
Running Buffer	20mM MOPS pH 7, 200mM NaCl, 5mM DTT

Table 26: List of buffers used for purification of His-TIR-TLR2

The supernatant is incubated with 2-3ml of Ni Sepharose® 6 Fast Flow (GE Healthcare) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins are washed by 25 CV of Washing Buffer (Table 26). The target protein is eluted with 20 ml Elution Buffer (Table 26). Eluted protein is concentrated (Vivaspin™, 3000 MWCO, GE Healthcare) to ~1ml and applied on equilibrated Superdex75 16/60 (GE Healthcare) column connected to Äkta Purifier (GE Healthcare). All fractions of affinity and size exclusion chromatography are analyzed on 15% SDS-PAGE. Fractions containing pure protein are concentrated, aliquoted, shock-frozen in liquid nitrogen and stored at -80°C.

3.2.3.3.7 Protein production and purification of NOD1 CARD domain

Domain boundaries for CARD domain of NOD1 were previously published (Srimathi, Robbins, Dubas, Seo, & Park, 2007) and cloned in pET15b derived vector with C-terminal His tag. Based on test expressions in *E. coli*, 3 hr expression in LB media at 37°C and 500 µM IPTG was the best condition. After 3 hr of expression at 37°C, cells are harvested in Sorvall RC 6+ centrifuge using F9S-4x1000y PTI rotor at 5000 rpm, 10 min. Pellet is resuspended in Lysis Buffer (Table 27) and the suspension is passed through a homogenizer (Avestin EmulsiFlex-C3) twice. The lysate is cleared by centrifugation at 16000 rpm, 45 min, 4°C.

Affinity Chromatography	
Buffers	**Composition**
Lysis Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME, Protease Inhibitor Cocktail tablet (Sigma), DNase, lysozyme, 0.5mM $MgSO_4$
Wash Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME, 20mM Imidazole
Elution Buffer	50mM HEPES pH 7.4, 300mM NaCl, 5% Glycerol, 5mM BME, 300mM Imidazole
Size Exclusion Chromatography	
Running Buffer	10mM HEPES pH 7.4, 100mM NaCl, 5mM DTT

Table 27: List of buffers used for purification of His-NOD1 CARD

The supernatant is incubated with 2-3ml of Ni Sepharose® 6 Fast Flow (GE Healthcare) equilibrated with Lysis Buffer, for 1 hr, 4°C with gentle stirring. Nonspecifically bound proteins are washed by 25 CV of Washing Buffer (Table 27). The target protein is eluted with 2x 20ml Elution Buffer (Table 27). Eluted protein is concentrated (Vivaspin™, 3000 MWCO, GE Healthcare) to ~1ml and applied on equilibrated Superdex75 16/60 (GE Healthcare) column connected to Äkta Purifier (GE Healthcare). All fractions of affinity and size exclusion chromatography are analyzed on 15% SDS-PAGE. Fractions containing pure protein are concentrated, aliquoted, and shock-frozen in liquid nitrogen and stored at -80°C.

3.2.4 Biochemical methods

3.2.4.1 SDS polyacrylamide gel electrophoresis (SDS PAGE)

SDS PAGE gels are used for quality control of purification protocols and for analysis of pull-down assays. The gels are composed of two parts: the upper stacking gel (5% acrylamide) and the lower resolving gel (12% or 15% acrylamide). The protocol is adapted from Laemmli (Laemmli 1970). Typically, samples are mixed with 2x DTT loading buffer, boiled at 95°C and ~10µl is loaded on the gel. SDS PAGE gels are run at constant current of 200V until the blue front reached the bottom of the gel. Following electrophoresis, gel is stained by boiling in SDS staining solution for 1-2 min in the microwave at maximum power, and destained by boiling in water for 5 min. Composition of buffers and solutions is listed in Section 3.1.11.3.

3.2.4.2 Western blot

For Western blotting, SDS PAGE gels are transferred onto a PVDF transfer membrane (Millipore) using a semi-dry blotting system (Trans-Bio® Turbo™ Transfer System) with Transfer Buffer. Membranes are incubated in the Blocking Buffer for 1 hr or overnight prior to 1 hr incubation with a primary antibody, and 1 hr incubation with secondary antibody. Membranes are developed with ECL reagent for HRP conjugated secondary antibodies or with BCIP (bromochloroindolyl phosphate) and NBT (nitro blue tetrazolium) for AP conjugated secondary antibodies. Electrogenerated chemiluminescence is visualized using Image Reader LAS 3000. Composition of buffers and solutions is listed in Section 3.1.11.3.

3.2.4.3 ThermoFluor assay

ThermoFlour assay allows screening the effects of different buffer conditions on stability of a protein (Ericsson *et al.* 2006; Cummings *et al.* 2006). Main benefit of ThermoFlour assay is determination of an optimal buffer in which a certain protein is stable. Prior to buffer test, protein and Sypro Orange dye (Sigma) concentrations are optimized. The best combination is added to desired buffers (Table 28) in a 96 wells plate (Multiplate 96-Well Unskirted PCR plate, BioRad), sealed and measured in a CFX96 Real-Time Thermal Cycler C1000 (BioRad). Data is analyzed using Bio-Rad CFX Manager software.

	1	2	3	4	5	6	7	8	9	10	11	12
A	50mM Acetate pH4.0	50mM Acetate pH4.4	50mM Citrate pH5.0	50mM Citrate pH5.4	50mM MES pH6.0	50mM MES pH6.4	50mM HEPES pH7.0	50mM HEPES pH7.4	50mM Tris-HCl pH8.0	50mM Tris-HCl pH8.4	50mM CHES pH9.0	50mM CHES pH9.4
B	50mM Acetate pH4.0 100mM NaCl	50mM Acetate pH4.4 100mM NaCl	50mM Citrate pH5.0 100mM NaCl	50mM Citrate pH5.4 100mM NaCl	50mM MES pH6.0 100mM NaCl	50mM MES pH6.4 100mM NaCl	50mM HEPES pH7.0 100mM NaCl	50mM HEPES pH7.4 100mM NaCl	50mM Tris-HCl pH8.0 100mM NaCl	50mM Tris-HCl pH8.4 100mM NaCl	50mM CHES pH9.0 100mM NaCl	50mM CHES pH9.4 100mM NaCl
C	50mM Acetate pH4.0 200mM NaCl	50mM Acetate pH4.4 200mM NaCl	50mM Citrate pH5.0 200mM NaCl	50mM Citrate pH5.4 200mM NaCl	50mM MES pH6.0 200mM NaCl	50mM MES pH6.4 200mM NaCl	50mM HEPES pH7.0 200mM NaCl	50mM HEPES pH7.4 200mM NaCl	50mM Tris-HCl pH8.0 200mM NaCl	50mM Tris-HCl pH8.4 200mM NaCl	50mM CHES pH9.0 200mM NaCl	50mM CHES pH9.4 200mM NaCl
D	50mM Acetate pH4.0 500mM NaCl	50mM Acetate pH4.4 500mM NaCl	50mM Citrate pH5.0 500mM NaCl	50mM Citrate pH5.4 500mM NaCl	50mM MES pH6.0 500mM NaCl	50mM MES pH6.4 500mM NaCl	50mM HEPES pH7.0 500mM NaCl	50mM HEPES pH7.4 500mM NaCl	50mM Tris-HCl pH8.0 500mM NaCl	50mM Tris-HCl pH8.4 500mM NaCl	50mM CHES pH9.0 500mM NaCl	50mM CHES pH9.4 500mM NaCl

Table 28: ThermoFlour Screen with buffers: salt concentration increases along the column, pH value increases along the row.

3.2.4.4 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is used to analyze overall secondary composition of protein. CD is based on differential absorption of left and right circularly polarized light by chiral molecules. Proteins composing of β-sheets or α-helices lead to different CD spectrum due to the different angles in the peptide bonds. Proteins structured into β-sheets will show a typical minimum at 218nm in the CD spectrum, while α-helical proteins show two minima, at 208 and 222nm. Based on this, CD spectroscopy can analyze secondary structure composition of proteins. Before the measurement, protein is dialyzed in a CD buffer (Section 3.1.11.3), which is also used for baseline correction. Samples are filled into Quartz cells (Hellma Analytics) with 1 mm path length. Spectra are recorded in the range of 190nm to 250nm at 25°C with a J-810 CD spectrometer (Jasco).

3.2.4.5 Dynamic light scattering

Dynamic light scattering (DLS) is used to determine size distribution profile of protein in solution. Sample preparation is done the same as for CD spectroscopy. The sample is additionally filtered through 0.2µm filter and filled into a dust-free DLS cuvette (Eppendorf). DLS measurements are carried out on Wyatt DynaPro temperature-controlled microsampler (Wyatt Technology Corporation), at 25°C. Measurement and analysis is done using DYNAMICS V6 software (Wyatt Technology Corporation).

3.2.4.6 MicroScale Thermophoresis (MST)

Macromolecules, such as proteins, migrate along a temperature gradient. Microscale thermophoresis (MST) takes advantage of this fact to study bimolecular interaction by monitoring changes in hydration shell, charge and size. The experimental procedure is according to the protocol from NanoTemper. Analysis is performed on Monolith NT.115. WD40-ATG16L1 protein is purified from insect cells and labelled according to the NanoTemper instructions. Dye used in NanoTemper kit binds to primary amino groups on solvent exposed lysines, with typically one lysine/protein being labeled. In this way labeling is random and, considering the average of solvent exposed lysines in a protein, the chance of dye interfering with binding is low. The concentration of the label protein is kept constant while the concentration of interaction partners varies. The signal is recorded for each condition. Any change in thermophoretic property is observed as a change in fluorescence intensity. Data is analyzed using MST software from NanoTemper.

3.2.5 Protein crystallization

3.2.5.1 Screening

Initial screening of crystallization conditions is done in 96-well sitting drop vapor diffusion Intelli® 93-3 LVR plates (Art Robbins Instruments, Hampton Research). Purified protein is concentrated and 0.2µl of protein, typically in the range from 5-10 mg/ml, was added to 0.2µl of reservoir solution. Volume of the reservoir is 800µl. Honeybee 961 crystallization robot (Zinsser Analytics) is used to dispense protein and reservoirs from commercially available screens (Table 15). Plates are covered with clear sealing tape to prevent evaporation and stored at 20°C or 4°C in an automated imaging robot from Formulatrix.

3.2.5.2 Refinement of initial crystals

Initial crystals were tested for diffraction at the home source (Table 29).

X ray home source	MX007 (Generator)	Rigaku
	RA-Micro 7 HFM (Anode)	
	VariMax HF (Optics)	
	Cryostream 700	Oxford Cryosystems

Table 29: In-house X-ray source used for initial testing of protein crystals

Initial protein crystals are reproduced and crystallization conditions are optimized using three strategies. Firstly, optimization is done by designing a grid and random screen around the initial condition using Formulatrix. In this way, pH and precipitant concentration are varied. In addition to incubation at 20°C and 4°C, plates are incubated at 30°C as well. Secondly, crystal growth promoting additives are tested using the Additive Screen (Hampton Research) as well as Silver Bullet Screen (Hampton Research). Seeding is used as a third method. Seeding is a useful method for obtaining higher quality crystals or new crystal forms. For seeding, initial crystals are crushed and used as seed stock. Seed stock is added to the drop containing protein and reservoir solution by OryxNano (Douglas Instruments Ltd). Plates are covered with clear sealing tape to prevent evaporation and stored at 20°C or 4°C in an automated imaging robot from Formulatrix.

3.2.5.3 Cryoprotection and flash cooling of protein crystals

When a protein crystal is exposed to X-rays, X-rays interact with electrons in the protein and introduce free radicals, especially when high energy synchrotron radiation is used. Freezing of crystals in liquid nitrogen and cooling them in the liquid nitrogen stream during exposure to X-rays reduces the free radical diffusion and minimizes crystal damage during data collection. However, most of damage is caused by not by protein but by water radicals which can diffuse and destroy protein crystal lattice. Additionally, freezing of aqueous solution leads to formation of crystalline ice, which may also damage the protein crystal due to expansion of the ice formation in the solvent s. Presence of crystalline ice also reduce the quality of data collection due to its inherent diffraction, which masks the diffraction from the protein crystals. Addition of a cryoprotectant, a substance used to form vitrous, amorphous ice during freezing, minimizes damage to the protein crystal, reduces diffusion of free radicals in the crystal during measurement and loss of diffraction due to radiation damage. To preserve the crystal form, high enough concentration of cryoprotectant is added to the original crystallization solution to obtain vitreous, amorphous ice. For flash cooling, crystals are rinsed in a fresh drop of mother liquor and then in the drop of cryo solution before being shock frozen in liquid nitrogen.

In this work, (2-R, 3-R)-butane-2,3-diol (Sigma) was successfully used as a cryoprotecting additive with a final concentration of 10 % (v/v).

3.2.5.4 Structure determination by X-ray crystallography

When exposed to an X-ray beam, a protein crystal diffracts a small percentage of X-rays in a certain direction and with certain intensities. Since X-rays are also waves, phase information is lost (and must be derived) due to physical detection of the diffracted photons. Thus, only intensity of the diffracted X-rays is recorded during the diffraction experiment whereas the phase information is lost, which is referred to as the phase problem. X-ray diffraction can be represented as reflection of X-rays on planes, as illustrated by Bragg's law (Figure 12).

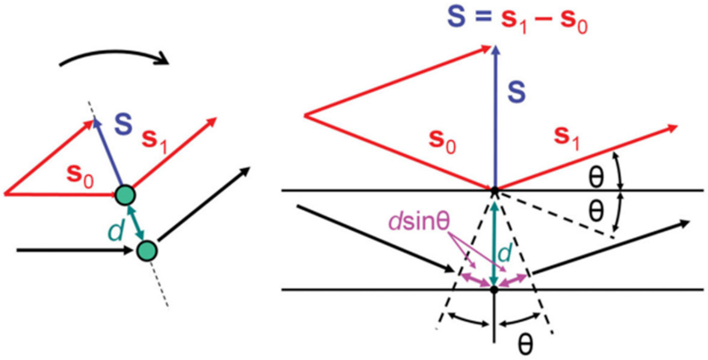


Figure 12: Graphical representation of Bragg's law. S is a scattering vector, given by the difference between incoming wave s0 and scattered wave s1. The second diagram has S pointing straight up. This depicts the graphical representation of Bragg's law and allows visualization of scattering waves as reflection on a set of imaginary parallel planes in a crystal. The distance between two planes is d. Scattering angle Θ is related to d and the path difference between (purple arrows) is 2d sinΘ. This difference must equal to nλ for constructive interference. (adapted from "Biomolecular Crystallography" Rupp 2010)

Bragg's law assumes that X-rays are reflected from of sets of (imaginary) Miller planes which are evenly distributed throughout the crystal. Reflected waves result in diffraction only when the condition of constructive interference is satisfied. This condition is referred to as Bragg's law and is represented by the following equation:

$$n\lambda = 2d\sin(\theta)$$

Where λ is the wavelength of incident wave, *d* is the spacing between the planes in the atomic lattice, and θ is the angle between the incident ray and the scattering planes.

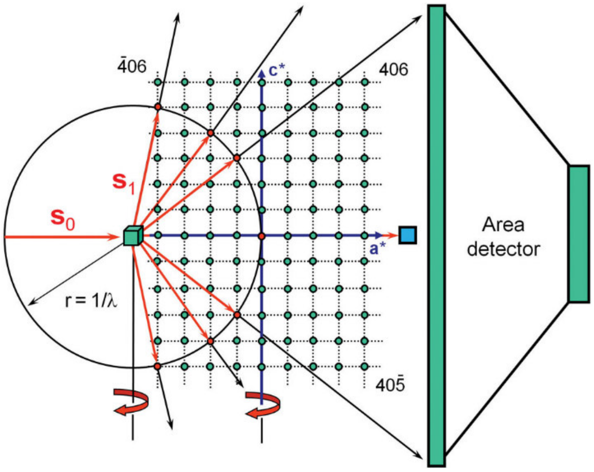


Figure 13: Geometrical construction of Ewald sphere. Bragg's diffraction condition can be visualized by Ewald sphere. The scattering vector S₀ is equal to 1/λ, which is the radius of the Ewald sphere. When the reciprocal lattice (green dots) is superimposed onto Ewald sphere, any point that crosses the Ewald sphere (red dots) will result in a reflection on the detector. (adapted from "Biomolecular Crystallography", (Rupp 2010)*)*

Conditions of positive interference resulting in diffraction can be represented by the Ewald sphere. Ewald sphere (Figure 13) is constructed with a radius of 1/λ around the X-ray beam. Green dots represent a reciprocal lattice, which is a set of imaginary points separated by a reciprocal vector. Reciprocal vector is easily generated based on two properties that define it: it is normal to Miller planes and the length of the vector is reciprocal to real space inter planar distance d. Any reciprocal lattice point that lies on the Ewald sphere will fulfill Bragg's diffraction condition and result in a reflection. Rotation of the crystal rotates the reciprocal lattice and brings other reciprocal lattice points to intersect with the Ewald sphere. Crystal rotation during data collection ensures higher completeness of the data set. A data set consists of intensities (amplitudes) of all recorded reflections.

X-ray diffraction data is collected at 100K at BESSY II synchrotron radiation source (Helmholtz Zentrum Berlin, Berlin, Germany). Data are processed with XDS software package (Kabsch 2010).

3.2.5.5 Experimental phasing and molecular replacement

Reflections that are recorded on the detector provide intensities (amplitudes) of complex structure factors F_{hkl}. This is necessary, but not sufficient for calculation of electron density and subsequent structure determination. Phase information is lost during physical measurement and has to be supplied from an external source to reconstruct the actual molecular structure from diffraction data.

As the phase angle cannot directly be determined, several indirect methods have been developed to overcome the phase problem. These methods include single/multiple isomorphous replacement (SIR/MIR), single/multiple wavelength anomalous dispersion (SAD/MAD) and molecular replacement (MR).

3.2.5.5.1 Experimental phasing

With no prior structural knowledge (ex. homologous structures) about the protein, SIR/MIR and SAD/MAD techniques are used for structure determination. For SIR/MIR, a heavy atom is introduced into the crystal (by soaking or co-crystallization). Heavy atoms provide the source of electronic differences in protein crystals. In order for SIR/MIR to be successful, the heavy atom must not alter the structure of protein molecules in the crystal, so the only difference between native and isomorphous (derivative) data set is the presence of the heavy atom. Depending how many derivatives (heavy atoms) are used, the method is referred to as single isomorphous replacement (SIR) or multiple isomorphous replacement (MIR).

Most heavy atoms, as well as other elements, are a source of anomalous signal. Due to the presence of anomalous signal at a certain wavelength (absorption edge of the element incorporated in the crystal), Friedel's law ($F_{(h,k,l)} = F_{(-h;-k,-l)}$) is not strictly true, which means the centrosymmetric reciprocal lattice points are no longer the same. This difference in intensity between two Friedel's mates is referred to as anomalous dispersion. By collecting data sets at different wavelengths (at the point of inflection, at the peak and at a high- and low- energy remote points) protein phases can be uniquely determined. In case of single wavelength anomalous dispersion (SAD) experiment, only peak data set is collected, where the anomalous difference is the largest. In case of crystal decay, the phases can still be calculated from the strong anomalous difference. For multiple wavelength anomalous dispersion (MAD) experiment, in addition to the data set at the peak wavelength, data is collected at the additional choices of wavelengths, in the following order: at the inflection, at low- and at high-energy. The main advantage of SAD/MAD method over SIR/MIR is, that in the case of SAD/MAD a single crystal is often enough to solve the structure.

3.2.5.5.2 Molecular replacement

In molecular replacement approach, initial phases are obtained from structurally similar models. Generally, a known structural model should have a sequence identity of ~30 % with the protein in question. Structures of suitable models are normally found via homology searches. The main challenge is to position the known model (search probe) in the unit cell (or asymmetric unit) of the target protein crystal. Intuitively, the search probe can be placed in the asymmetric unit and three positional coordinates and three angles at each grid point are varied until the match is found. This is computationally demanding and can be broken down into two 3-dimensional searches: rotational search determines the correct orientation of the search probe, followed by translational search which narrows down its exact position in the unit cell. Once the correct positon is determined, the Fourier synthesis of the molecule provides calculated phases and calculated intensities of the search model.

Fourier transform, a reciprocal transform of periodic functions between two domains, is used to calculate electron density of the real space structure from experimentally obtained complex structure factors F_{hkl}. Electron density equation

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right|e^{-2\pi\cdot i(hx+ky+lz-\alpha_{hkl})}$$

demonstrates that electron density (ρ) is a Fourier summation (synthesis) with all complex structure factor amplitudes and the phases (α_{hkl}) over the entire reciprocal space ($\frac{1}{V}$ is volume of the reciprocal lattice). This allows quick transform from real to reciprocal space, and the other way around.

The calculated phases derived from the molecular replacement model are combined with measured intensities from the target protein crystal and used in Fourier transform, resulting in interpretable electron density map.

3.2.5.6 Model building and refinement

After initial phase information has been obtained by molecular replacement, the model of the protein can be built. Since initial phases still contain some errors, they have to be improved and refined. This is a reiterative process: improved phases lead to an improved map, with more details, resulting in a more detailed and complete model, which is then used to calculate new phases and leads to an even better map. During this process, two key factors to observe are observed: R_{work} and R_{free}. The first value, R_{work} is a measure of the quality of fit of the built model and the experimental diffraction data. During refinement, the value of R_{work} should improve and eventually reach 15-25%, depending on the quality of the data. The greatest danger with R_{work} is overfitting. Overfitting is loosely defined as building in arbitrary atoms or allowing chemically unfavorable position to reduce R_{work} and have a better fit. The second R value, R_{free} is introduced to avoid this problem. R_{free} compares the small subset of the data (usually 5-10 %) not included in R_{work} calculation and is thus independent. The R_{free} is always a few percent higher than the R_{work}.

In addition to R factors, average *B* value, or temperature factor, indicates precision of the atom positions. Disorder of the crystal, or flexibility of the side chain can lead to uncertainty in the exact atomic position. This uncertainty is represented with high *B* factors. However, since exact position of atoms cannot be directly observed, B factors are a result of fitting of a model to the diffraction data. Thus, *B* factors are partially dependent on the model used for phase calculation and refinement process and cannot be used as an absolute value in direct comparison between different structures.

The structural model of ATG16L1_G303_E354AK355A_P607Y was built using COOT (Emsley and Cowtan 2004) and refinement was done in Phenix (Adams *et al.* 2010).

4 Results

4.1 Characterization and structure determination of C-terminal domain of ATG16L1

4.1.1 Generation of constructs for crystallization

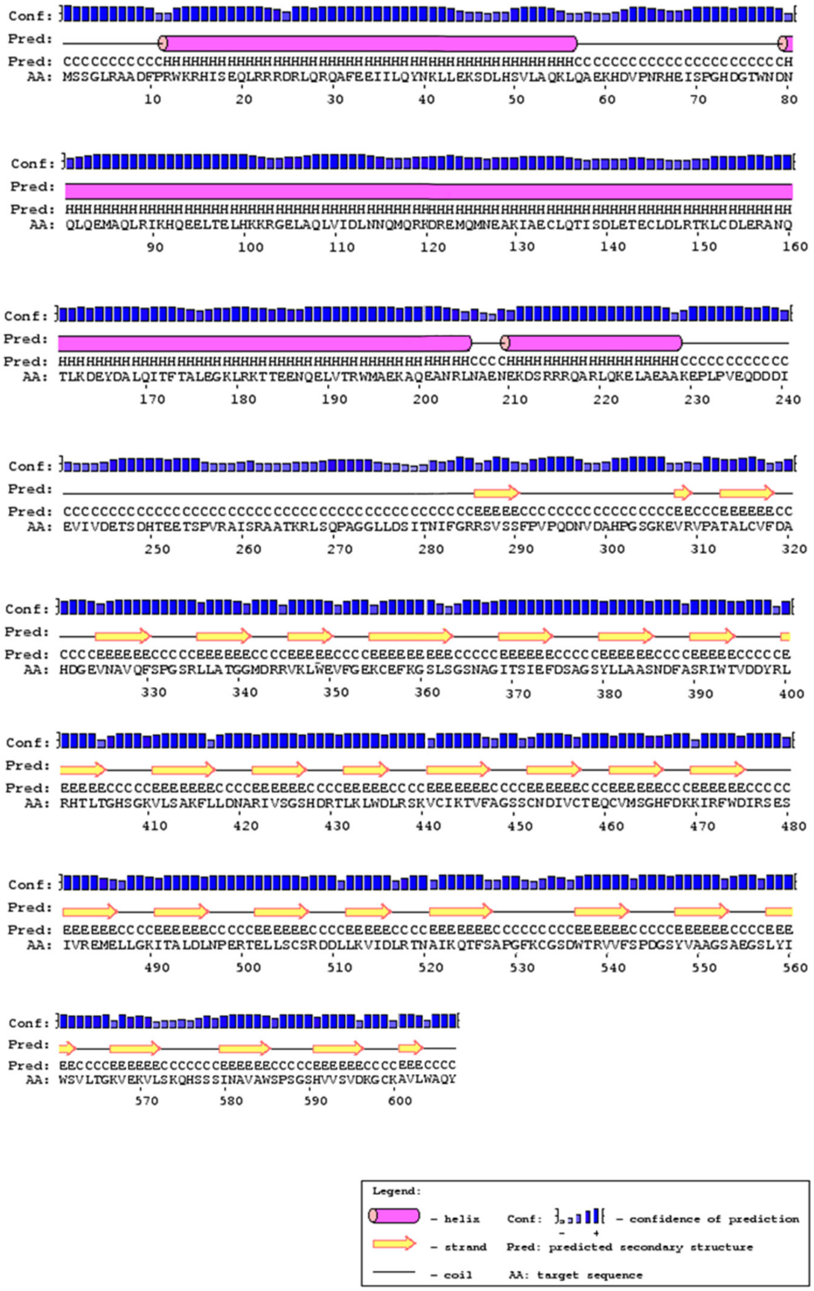
A main goal of this project was structure determination of the C-terminal domain of human ATG16L1 by X-ray diffraction analysis of protein crystals. Structural studies of any protein can be roughly divided into 8 main stages: bioinformatic analysis and construct design, cloning, searching for expression conditions, large scale purification, crystal growth, crystal diffraction, data collection and processing and finally structure determination and model building. Growing crystals of a desired protein is a crucial step in this process and all the efforts are made to make protein crystal packing possible. However, crystallizing a protein is a highly empirical process whose success cannot be predicted or guaranteed. Few of many possible obstacles on the gene to structure endeavor will be addressed in the following text.

4.1.1.1 Human ATG16L1

4.1.1.1.1 Bioinformatic analysis

Each crystallographic study starts with gathering of information about the gene, the protein it encodes, and what can be assumed about it, based on related proteins or folds. Ineluctable requirement for any macromolecular structure determination is primary sequence of the protein. UniProt (http//www.uniprot.org) is a reliable database with information about protein sequences and annotated data. In UniProt, human ATG16L1 is classified under identifier Q676U5.

The primary peptide sequence is used to calculate secondary structures of proteins based on amino acid sequence composition, taking into account homology models and proteins with similar domains or folds. PSIPRED v3.3 (UCL Department of Computer Science, Bioinformatics Group, http://bioinf.cs.ucl.ac.uk/psipred/) derives its prediction from PSI-BLAST and provides an output file which can be used as a guideline for construct design. PSIPRED analysis of human ATG16L1 is shown in Figure 14. Based on the prediction, structured, mainly helical N-terminal region of human ATG16L1 ends by residue 230, followed by a random coil region (230-284). First predicted β-strand starts at position 284, and the rest of the protein (284-607) is structured in regularly distributed β-strands.



```
Conf:
Pred:
Pred: CCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCH
  AA: MSSGLRAADFPRWKRHISEQLRRRDRLQRQAFEEIILQYNKLLEKSDLHSVLAQKLQAEKHDVPNRHEISPGHDGTWNDN
               10        20        30        40        50        60        70        80

Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
  AA: QLQEMAQLRIKHQEELTELHKKRGELAQLVIDLNNQMQRKDREMQMNEAKIAECLQTISDLETECLDLRTKLCDLERANQ
               90       100       110       120       130       140       150       160

Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCC
  AA: TLKDEYDALQITFTALEGKLRKTTEENQELVTRWMAEKAQEANRLNAENEKDSRRRQARLQKELAEAAKEPLPVEQDDDI
              170       180       190       200       210       220       230       240

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEECCCCCCCCCCCCCCCCCEECCCEEEEEECC
  AA: EVIVDETSDHTEETSPVRAISRAATKRLSQPAGGLLDSITNIFGRRSVSSFPVPQDNVDAHPGSGKEVRVPATALCVFDA
              250       260       270       280       290       300       310       320

Conf:
Pred:
Pred: CCCCEEEEEECCCCCEEEEEECCCCEEEEECCCCEEEEEEEEECCCCCEEEEEECCCCCCEEEEEECCCCEEEEECCCCE
  AA: HDGEVNAVQFSPGSRLLATGGMDRRVKLWEVFGEKCEFKGSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRL
              330       340       350       360       370       380       390       400

Conf:
Pred:
Pred: EEEEECCCCCEEEEEEECCCCEEEEEECCCCEEEEECCCCEEEEEEECCCCEEEEEECCCEEEEEECCCEEEEEECCCCC
  AA: RHTLTGHSGKVLSAKFLLDNARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSES
              410       420       430       440       450       460       470       480

Conf:
Pred:
Pred: EEEEEECCCCEEEEEECCCCCEEEEEECCCCEEEEECCCCEEEEEEECCCCCCCCCEEEEEECCCCCEEEEEECCCCEEE
  AA: IVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAIKQTFSAPGFKCGSDWTRVVFSPDGSYVAAGSAEGSLYI
              490       500       510       520       530       540       550       560

Conf:
Pred:
Pred: EECCCCEEEEEECCCCCCCEEEEEECCCCCEEEEEECCCCEEECCCC
  AA: WSVLTGKVEKVLSKQHSSSINAVAWSPSGSHVVSVDKGCKAVLWAQY
              570       580       590       600
```

Figure 14: The secondary structure annotation of the human full length ATG16L1 amino acid sequence by PSIPRED v3.3 Pink cylinders indicate α-helices, flat black lines indicate coil, and yellow arrows represent β-strands.

For prediction of tertiary structure based on amino acid sequence, Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) compares the submitted sequence with similar proteins with known structures and predicts three dimensional fold of the protein. The result from Phyre2 of C-terminal domain of human ATG16L1 is shown in Figure 15: Phyre2 model of C-terminal domain of human ATG16L1. Based on the Phyre2 prediction, C-terminal domain of human ATG16L1 folds into a WD40 propeller domain, with high confidence.

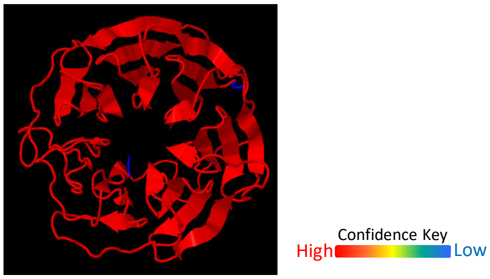


Figure 15: Phyre2 model of C-terminal domain of human ATG16L1

4.1.1.1.2 Construct design

Since protein crystallization is a rather empirical process and small changes on protein level can have profound impact on protein solubility and crystallization, identifying a suitable construct requires parallel screening of multiple protein constructs. Protein construct can be modified either on DNA or on protein level. First major bottleneck of crystallography is engineering of a DNA construct encoding a soluble domain. Construct design consists of manipulating protein on nucleic acid level to yield soluble, compact and stable protein prone to crystal packing. Since many details can be modified on a DNA level, it is more economical and practical to design and test multiple constructs in parallel at the beginning of the project, including orthologous proteins (discussed later), in hopes that one of them will yield a soluble protein that will self-assemble into diffracting crystals.

Naturally, a structure of a full length human ATG16L1 (M1-Y607) would provide the greatest deal of information about the protein. However, due to its size and multiple domain architecture, it might be difficult to express, purify and crystallize it. As a rule of thumb, single, compact domains are considered to be soluble, to able to fold independently and more likely to arrange themselves in a way that enables formation of the few weak crystal contacts. Thus, crystallographic studies of the C-terminal domain of human ATG16L1 was defined as a main aim of this project. PSIPRED prediction of secondary structure elements suggest the start of the first β-sheet of C-terminal domain of human ATG16L1 is at position G284, which was previously chosen as the start for one of the constructs. Prediction servers are based on sequence similarities between query and already known structures and must be taken with some reservations. The first predicted β-sheet of the C-terminal domain of human ATG16L1 could thus be a statistical artifact or not part of the C-terminal domain. Based on chemical and physical properties of amino acids following R290, the predicted end of the first β-sheet, G303 was chosen as another construct boundary previously designed in our lab

Since Phyre2 predicted the C-terminal domain to fold into WD-40 domain. These domains often have additional adjacent secondary functionally important structural elements (or entire domains). In the case of DDB2, a WD-40 domain involved in UV DNA damage recognition, a helix-loop-helix motif positioned N-terminally to the WD-40 domain is important for its interaction with DDB1 (Scrima *et al.*, 2008). With this in mind, K266 was chosen as a start of another C-terminal ATG16L1 construct, in order not to omit potentially functional elements.

Figure 16 shows a graphical representation of the constructs of human ATG16L1 used in this study.

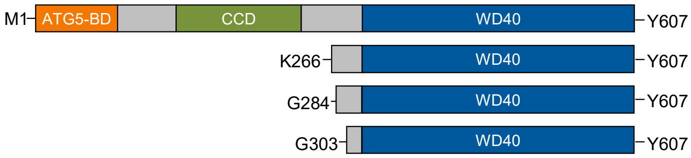


Figure 16: Construct boundaries of C-terminal domain of human ATG16L1 based on bioinformatics predictions

4.1.1.2 Mouse ATG16L1

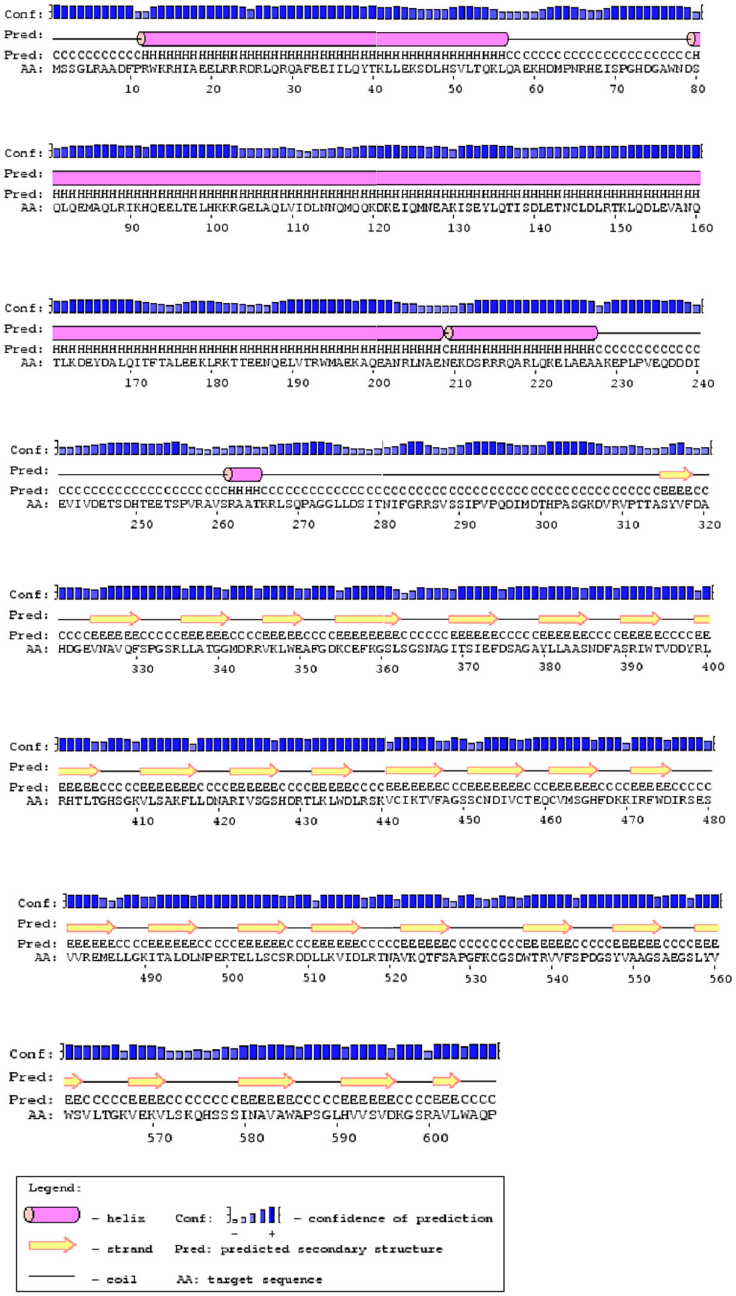
As previously mentioned, protein's tendencies towards crystal formation cannot be predicted. Often, small adjustments on the amino acid sequence level can prove beneficial and tilt the balance towards crystal assembly. Apart from designing different constructs of the same protein, orthologues of target protein might be more likely to crystalize. Orthologues often retain the same function, thus ensuring conservation of functionally important, core residues, while variation between orthologues comes from differences in non-essential, surface-exposed residues. Thus, structural biologists often take advantage of construct engineering performed by nature by testing orthologues of the target protein.

```
hsATG16L1      MSSGLRAADFPRWKRHISEQLRRRDRLQRQAFEEIILQYNKLLEKSDLHSVLAQKLQAEK      60
mmATG16L1      MSSGLRAADFPRWKRHIAEELRRRDRLQRQAFEEIILQYTKLLEKSDLHSVLTQKLQAEK      60
               *****************:*:*******************.************:*******

hsATG16L1      HDVPNRHEISPGHDGTWNDNQLQEMAQLRIKHQEELTELHKKRGELAQLVIDLNNQMQRK      120
mmATG16L1      HDMPNRHEISPGHDGAWNDSQLQEMAQLRIKHQEELTELHKKRGELAQLVIDLNNQMQQK      120
               **:************:***.**************************************:*

hsATG16L1      DREMQMNEAKIAECLQTISDLETECLDLRTKLCDLERANQTLKDEYDALQITFTALEGKL      180
mmATG16L1      DKEIQMNEAKISEYLQTISDLETNCLDLRTKLQDLEVANQTLKDEYDALQITFTALEEKL      180
               *:*:*******:* *********:******** *** ******************** **

hsATG16L1      RKTTEENQELVTRWMAEKAQEANRLNAENEKDSRRRQARLQKELAEAAKEPLPVEQDDDI      240
mmATG16L1      RKTTEENQELVTRWMAEKAQEANRLNAENEKDSRRRQARLQKELAEAAKEPLPVEQDDDI      240
               ************************************************************

hsATG16L1      EVIVDETSDHTEETSPVRAISRAATKRLSQPAGGLLDSITNIFGRRSVSSFPVPQDNVDT      300
mmATG16L1      EVIVDETSDHTEETSPVRAVSRAATKRLSQPAGGLLDSITNIFGRRSVSSIPVPQDIMDT      300
               *******************:******************************:***** :**

hsATG16L1      HPGSGKEVRVPATALCVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEVFGEKCEFKG      360
mmATG16L1      HPASGKDVRVPTTASYVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEAFGDKCEFKG      360
               **.***:****:**  **********************************.**:******

hsATG16L1      SLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDN      420
mmATG16L1      SLSGSNAGITSIEFDSAGAYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDN      420
               ******************:*****************************************

hsATG16L1      ARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSES      480
mmATG16L1      ARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSES      480
               ************************************************************

hsATG16L1      IVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAIKQTFSAPGFKCGSDWTRV      540
mmATG16L1      VVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAVKQTFSAPGFKCGSDWTRV      540
               :****************************************:******************

hsATG16L1      VFSPDGSYVAAGSAEGSLYIWSVLTGKVEKVLSKQHSSSINAVAWSPSGSHVVSVDKGCK      600
mmATG16L1      VFSPDGSYVAAGSAEGSLYVWSVLTGKVEKVLSKQHSSSINAVAWAPSGLHVVSVDKGSR      600
               *******************:*************************:*** ********.:

hsATG16L1      AVLWAQY   607
mmATG16L1      AVLWAQP   607
               ******
```

Figure 17: Amino acid sequence alignment between human ATG16L1 (hsATG16L1, UniProt ID Q676U5) and mouse ATG16L1 (mmATG16L1, UniProt ID Q8C0J2) by Clustal Omega.

Closely related species are likely to display very similar sequences between orthologues. Mouse (*Mus Musculus*) ATG16L1 is highly similar to human ATG16L1 (94 % protein sequence identity and 96 % protein sequence similarity) as shown by amino acid sequence alignment by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Figure 17).

Based on secondary structure prediction of mouse ATG16L1 (Figure 18), overall architecture is very similar to human ATG16L1. N-terminal, mostly α-helical part is predicted to be until position 288, followed by largely random coil region (288-314).



```
Conf:
Pred:
Pred: CCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCH
  AA: MSSGLRAADFPRWKRHIAEELRRRDRLQRQAFEEIILQYTKLLEKSDLHSVLTQKLQAEKHDMPNRHEISPGHDGAWNDS
              10        20        30        40        50        60        70        80

Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
  AA: QLQEMAQLRIKHQEELTELHKKRGELAQLVIDLNNQMQQKDKEIQMNEAKISEYLQTISDLETNCLDLRTKLQDLEVANQ
              90       100       110       120       130       140       150       160

Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCC
  AA: TLKDEYDALQITFTALEEKLRKTTEENQELVTRWMAEKAQEANRLNAENEKDSRRRQARLQKELAEAAKEPLPVEQDDDI
             170       180       190       200       210       220       230       240

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEECC
  AA: EVIVDETSDHTEETSPVRAVSRAATKRLSQPAGGLLDSITNIFGRRSVSSIPVPQDIMDTHPASGKDVRVPTTASYVFDA
             250       260       270       280       290       300       310       320

Conf:
Pred:
Pred: CCCCEEEEEECCCCCEEEEEECCCCEEEEECCCCEEEEEEEECCCCCCEEEEEECCCCCEEEEEECCCCEEEEECCCCEE
  AA: HDGEVNAVQFSPGSRLLATGGMDRRVKLWEAFGDKCEFKGSLSGSNAGITSIEFDSAGAYLLAASNDFASRIWTVDDYRL
             330       340       350       360       370       380       390       400

Conf:
Pred:
Pred: EEEEECCCCCEEEEEEECCCCEEEEEECCCCEEEEECCCCEEEEEEECCCEEEEEEECCCEEEEEECCCCEEEEECCCCC
  AA: RHTLTGHSGKVLSAKFLLDNARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSES
             410       420       430       440       450       460       470       480

Conf:
Pred:
Pred: EEEEEECCCCEEEEEECCCCCEEEEEECCCEEEEEECCCCCEEEEEECCCCCCCCCEEEEEECCCCCEEEEEECCCCEEE
  AA: VVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAVKQTFSAPGFKCGSDWTRVVFSPDGSYVAAGSAEGSLYV
             490       500       510       520       530       540       550       560

Conf:
Pred:
Pred: EECCCCCEEEECCCCCCCCEEEEEECCCCCEEEEEECCCCEEECCCC
  AA: WSVLTGKVEKVLSKQHSSSINAVAWAPSGLHVVSVDKGSRAVLWAQP
               570       580       590       600
```

Figure 18: PSIPRED v3.3 output of mouse full length ATG16L1 amino acid sequence

In the case of mouse ATG16L1, the first putative N-terminal β-strand of the WD-40 domain starts at position 315 with serine. A short valine-proline patch upstream in the sequence seems like a site with least potential for disturbing the overall fold. Thus, proline 311 was chosen as the start of one of the C-terminal construct for mouse ATG16L1. Following the same logic as in designing human ATG16L1 construct, mouse construct starting at A320 skips the first β-strand, which is predicted to be short and with lower confidence than the rest of the domain, and construct starting at P271 might include a functionally important preceding sequence motif.

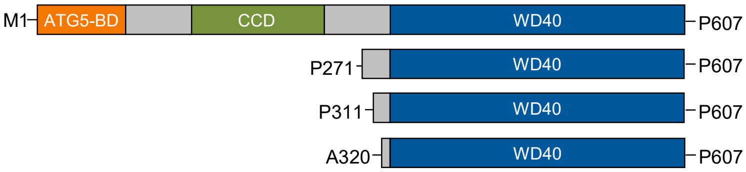


Figure 19: Construct boundaries of C-terminal domain of mouse ATG16L1 based on bioinformatics predictions

All mouse ATG16L1 constructs (Figure 19) are cloned directly into pFastbacDual vector with Strep and His tags, for virus production and expression in insect cells.

4.1.2 Expression and purification of human ATG16L1 constructs

Most proteins used for crystallization studies are recombinantly overexpressed from heterologous hosts. Many different expression systems are available, each with its advantages and disadvantages. Two most commonly used, which were also used in this work, are *Escherichia coli* (*E. coli*) and insect cells. *E. coli* is the most prevalent bacterial expression system. Due to a wide range of cloning and expression tools, *E. coli* provide a very versatile expression platform. Small scale test expression of all the constructs of human ATG16L1 was previously screened in *E. coli* but did not yield soluble protein production for any of the ATG16L1 constructs.

The second most common expression platform for proteins is a eukaryotic system based on insect cells and baculoviruses. Baculoviruses infect insect cells and viral genome has been manipulated to enable large scale recombinant protein production. In the baculovirus expression vector system (BEVS), naturally occurring polyhedron gene, produced during very late stage of viral replication, is under control of a very strong promoter. For *in vitro* protein production, polyhedrin gene is replaced with DNA coding for the recombinant protein of interest by homologous recombination. Virus is produced and amplified in Sf-21 cells. At P2 stage, Sf-21 cells are tested for protein production. If protein is soluble and over-expressed, P3 virus is used for large scale protein production from High Five cells. More than three passages of the virus are not recommended, due to high mutation rate of their genomes and their selection for infectious viruses, which do not necessarily encode the gene of interest. Apart from higher cost and longer time needed for protein production, variability between different batches is high, due to both virus and insect cells. Amount of virus used for protein production and harvest time has to be optimized for each virus, since protein expression depends on culture and infection conditions. Viral titer varies with each production cycle, and cell culture passage number and rate of division play an important role in the infection process. Additionally, stability of the virus is different depending on the gene of interest it encodes. The reason for this is not well understood and stability of the virus cannot be predicted but has to be experimentally tested. Some viruses which can be stored in the media with 5% FCS at 4°C with no detectable reduction in quality of protein production. Such are the viruses carrying the gene for C-terminal domain of human

ATG16L1. On the other hand, viruses used for expression of full length human ATG16L1 must be used up within days of production. To circumvent this problem, BIIC (baculovirus infected insect cells) with virus coding for full length human ATG16L1 were produced. In this way, also the variability in virus quality and titer is under better control.

Insect cells are quite tedious to maintain under optimal conditions for infection. If the cells divide too fast, they might grow too quickly and enter lag phase before virus can infect enough of the cells for successful protein production. In case of cell passage number being too low, or too high, cells might not be as receptive for the virus uptake. This is also the case when the cells reach a cell density above their usual range (above 3-4x10^6 cells/ml). Thus, much time is invested in keeping both insect cells as well as the virus under optimal conditions, in order for protein production to be effective.

4.1.3 Expression and purification of full length ATG16L1

4.1.3.1 Strep tagged full length human ATG16L1

Since proteins are expressed in cells, they are cloned with a tag, which assists in separation of target proteins from other cellular content. In the case of full length human ATG16L1, due to its size, only small tags were added on the N-terminal end. Initially, human ATG16L1 was cloned with N-terminal His and Strep tag, separated from the protein by a TEV cleavage site, and expression was tested in insect cells. Full length human ATG16L1 with a small tag is approximately 70 kDa in size. Protein with either tag was expressed (Figure 20) but the Strep tagged construct resulted higher yield and purity. A band corresponding to a 70 kDa protein is observed in Strep pull down from Sf-21 cells infected with P2 Strep ATG16L1 virus. This band has been identified as human ATG16L1 by tryptic digest and mass spectrometry (sequence coverage 38 %, score for human ATG16L1 of 226, next best score for an unnamed protein of 54). Apart from mass spectrometry, presence of human ATG16L1 was confirmed by Western blot against ATG16L1. The two upper bands in the Strep pull down (marked with asterisk (*) in Figure 20) are native to insect cells and non-specifically bind the Strep beads. The band bellow 14 kDa marker is Streptavidin from the beads (Figure 20).

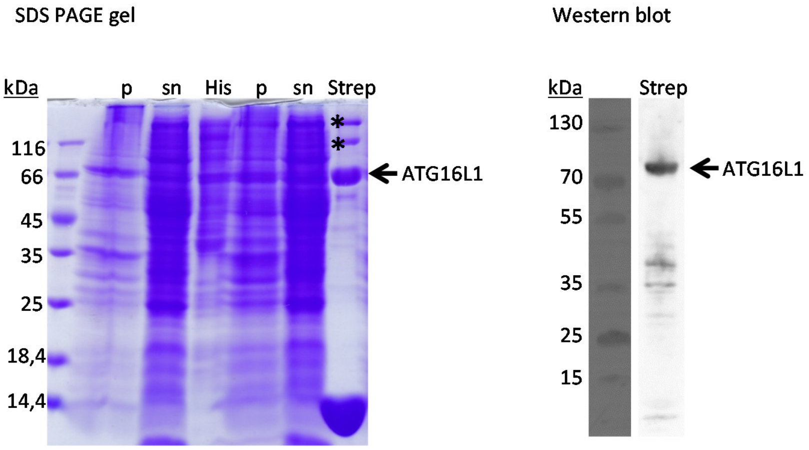


Figure 20: Small scale test expression of His tagged and Strep tagged human full length ATG16L1 from Sf-21 cells, 3 days post infection. (Legend: p = pellet, sn = supernatant, His = Ni-NTA beads, Strep = Strep-Tactin Macroprep beads) Left figure: SDS-PAGE analysis of small scale test expression. Right figure: Western Blot on the beads of Step tagged human full length ATG16L1 from Sf-21 cells, 3 days post infection. Primary antibody: anti-ATG16L1 (Sigma, A7356) 1:1000, Secondary antibody: Peroxidase conjugated AffiniPure Goat Ant-Rabbit IgG (1:2500). Detection: Lumi-Light Western Blotting Substrate (Roche). Bands marked with asterisk (*) is an insect cell protein unspecific binding Strep beads.

First large scale production of Strep tagged human full length ATG16L1 was done in Sf-21 cells. Following gravity affinity chromatography on Strep-Tactin Macroprep column, protein was concentrated and loaded on Superdex 200 16/60 size exclusion column. Fractions were analyzed on 12 % SDS PAGE gel (Figure 21A).

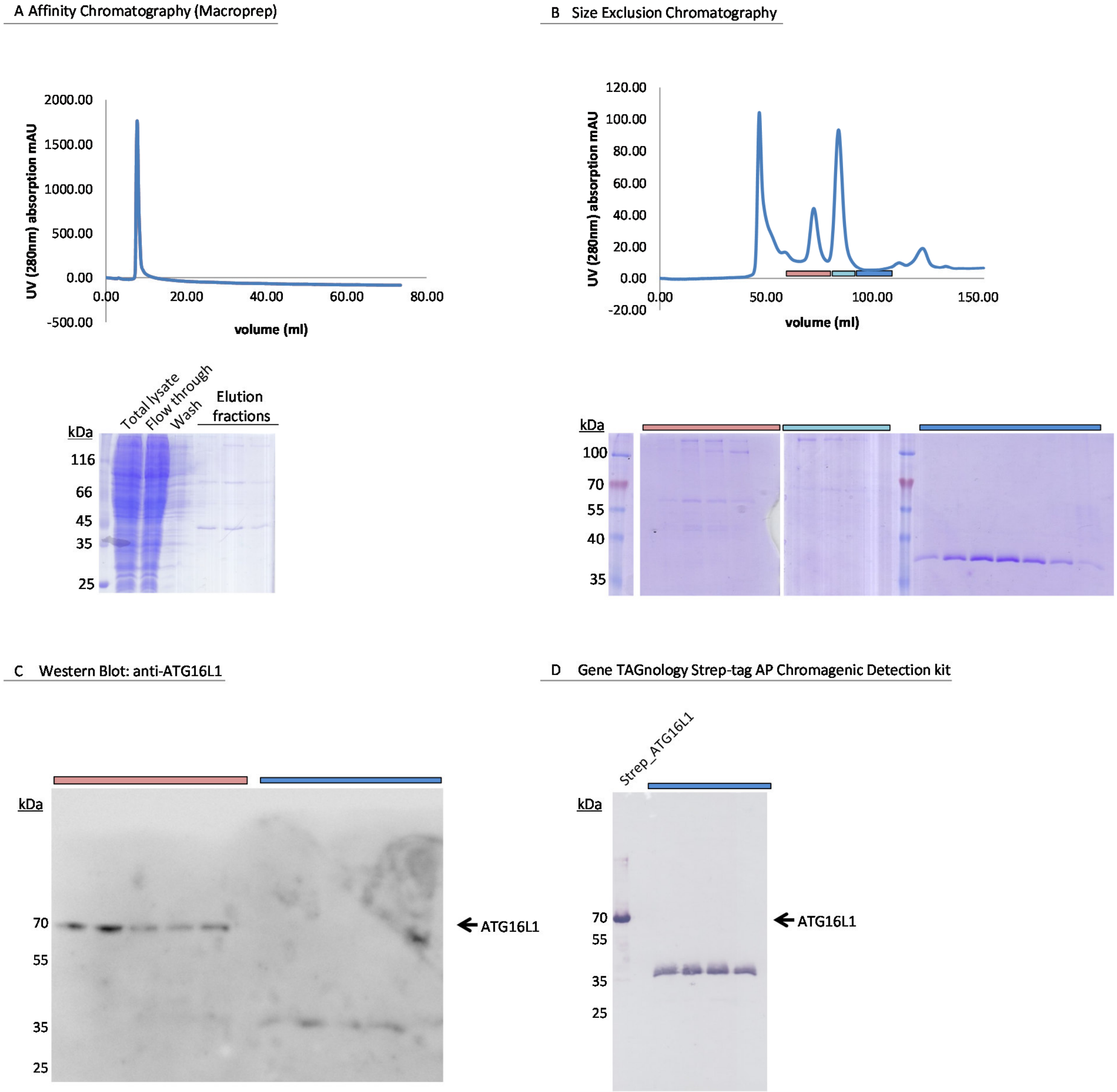


Figure 21: Large scale purification of Strep tagged full length human ATG16L1. A) 2 L culture of Sf-21 cells (0.8x10^6 cells/ml) infected with P3 Strep tagged full length human ATG16L1 virus. Cells were harvested 3 days post infection. Cleared lysate was loaded on Strep-Tactin Macroprep 10 ml resin connected to Äkta Purifier system, thoroughly washed and eluted with 2.5 mM desthiobiotin (elution profile shown in the right panel) in 10 ml elution fractions. B) Elution fractions from affinity chromatography were combined, concentrated in Vivaspin 10000 MWCO, and loaded on S200 16/60 size exclusion chromatography. Each fraction was analyzed on 12 % SDS-PAGE gel. C) Western blot against ATG16L1. Primary antibody: anti-ATG16L1 (Sigma, A7356) 1:1000, Secondary antibody: Peroxidase conjugated AffiniPure Goat Ant-Rabbit IgG (1:2500). Detection: Lumi-Light Western Blotting Substrate (Roche). D) Gene TAGnology Strep-tag AP Chromagenic Detection kit (iba BioTAGnology) with elution fraction after gel filtration of Strep full length human ATG16L1. Control: sample from test expression (Strep_ATG16L1).

The most prominent band from fractions D11-D5 did not correspond to the size of the full length ATG16L1 (70 kDa) but to a protein of a molecular weight of approx. 37 kDa. However, since it is the main band observed (Figure 21B) it was sent for mass spectrometry identification. Tryptic digest and mass spectrometry of peptides was not able to find a hit in the human database. Since affinity purification using Step tag is generally quite specific, elution fractions were submitted to anti-ATG16L1 Western blot (Figure 21C) and a Gene TAGnology Strep-tag AP Chromagenic Detection kit (iba BioTAGnology), which detects Strep tag on the principle of a Western Blot (Figure 21D). As a control for Strep-tag Detection kit, samples from small scale test expression of Step tagged full length human ATG16L1, (already confirmed as human ATG16L1 by mass spectrometry and anti-ATG16L1 Western blot), were used. From the Western blot analysis of size elution fractions, full length ATG16L1 is in the void volume, indicating a large soluble aggregate. A weak signal for ATG16L1 is present in the fractions marked with a blue bar (Figure 21C). However, Strep-tag detection kit confirmed the presence of the Strep tag on the protein in the fractions marked with a blue bar (Figure 21D), which would indicate that it is a degradation product of the full length ATG16L1. The tag was fused to the N-terminal end of the protein, and is still present in the degradation product. Based on this, it can be assumed that the N-terminal part is intact. To identify C-terminal boundary of the degradation product, it was sent for full mass determination. Measured mass of the sample was 38108 Da. Primary amino acid sequence of full length ATG16L1 including the sequence for the tag and the TEV cleavage site was used to calculate that C-terminal residue of the degradation product is E307. Based on the secondary structure prediction of the full length human ATG16L1 (Figure 14), E307 residue is located at the beginning of the C-terminal domain, predicted to be a WD-40 (Figure 15Figure 15). Since position 307 is close to 303, which was already used as a construct boundary (Figure 16), no additional constructs were generated.

Large scale production of full length human ATG16L1 proved to be a major challenge. Due to low yield from Sf-21 cells, large scale production was attempted in HighFIve cells. Since the virus was unstable, BIIC stock was used for infecting large scale production in HighFive cells. Protein expression was confirmed by measuring yellow fluoresce by Guava EasyCyte Mini (Merck Milipore) and small scale test expression protocol on 10 ml of 1 L culture. The protein was soluble in test expression, when beads used for binding to the protein were boiled and loaded on SDS PAGE gel. However, when elution of the bound protein was attempted in large scale production, initially it appeared that the protein is not expressed, or only at a very low level (Figure 22).

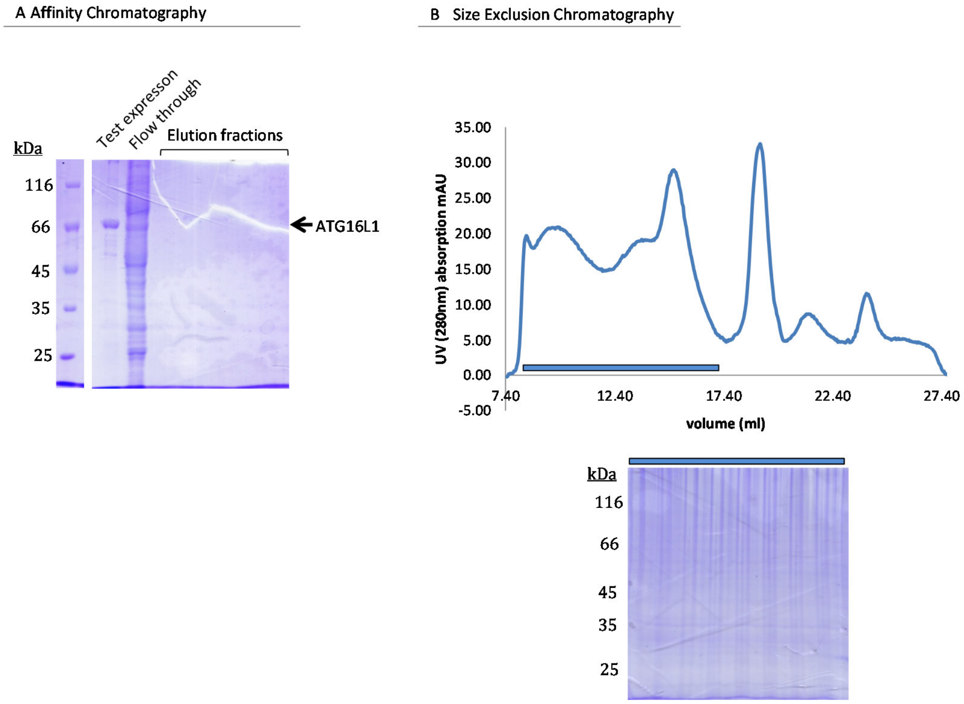


Figure 22: Large scale purification of Strep tagged full length human ATG16L1. A) Affinity chromatography was done by gravity flow, using Step High Capacity beads. 10 ml of the 1 L culture were used for test expression and beads were loaded as a control of protein expression (Test expression lane). Protein is eluted in 5x 5 ml fractions with 5 mM desthiobiotin. B) First three fractions from affinity chromatography were concentrated and loaded on S200 10/30 connected to Äkta Purifier system. Elution fractions of gel filtration run were analyzed on 12 % SDS PAGE gel.

Purification of Strep tagged full length human ATG16L1 was attempted many times, with no improvement of the yield. Protein production was always observed from 10 ml small test expression, but large scale purification showed no soluble protein. To optimize and enhance the success of the experiment, different culture volumes and flask diameters for growing insect cell cultures were extensively tested, with no detected improvement in protein production. Furthermore, in addition to High Capacity beads used for the purification in Figure 22, elution was tested from two additional resins used for Strep affinity: Sepharose (Figure 23A) and Macroprep (Figure 23B).

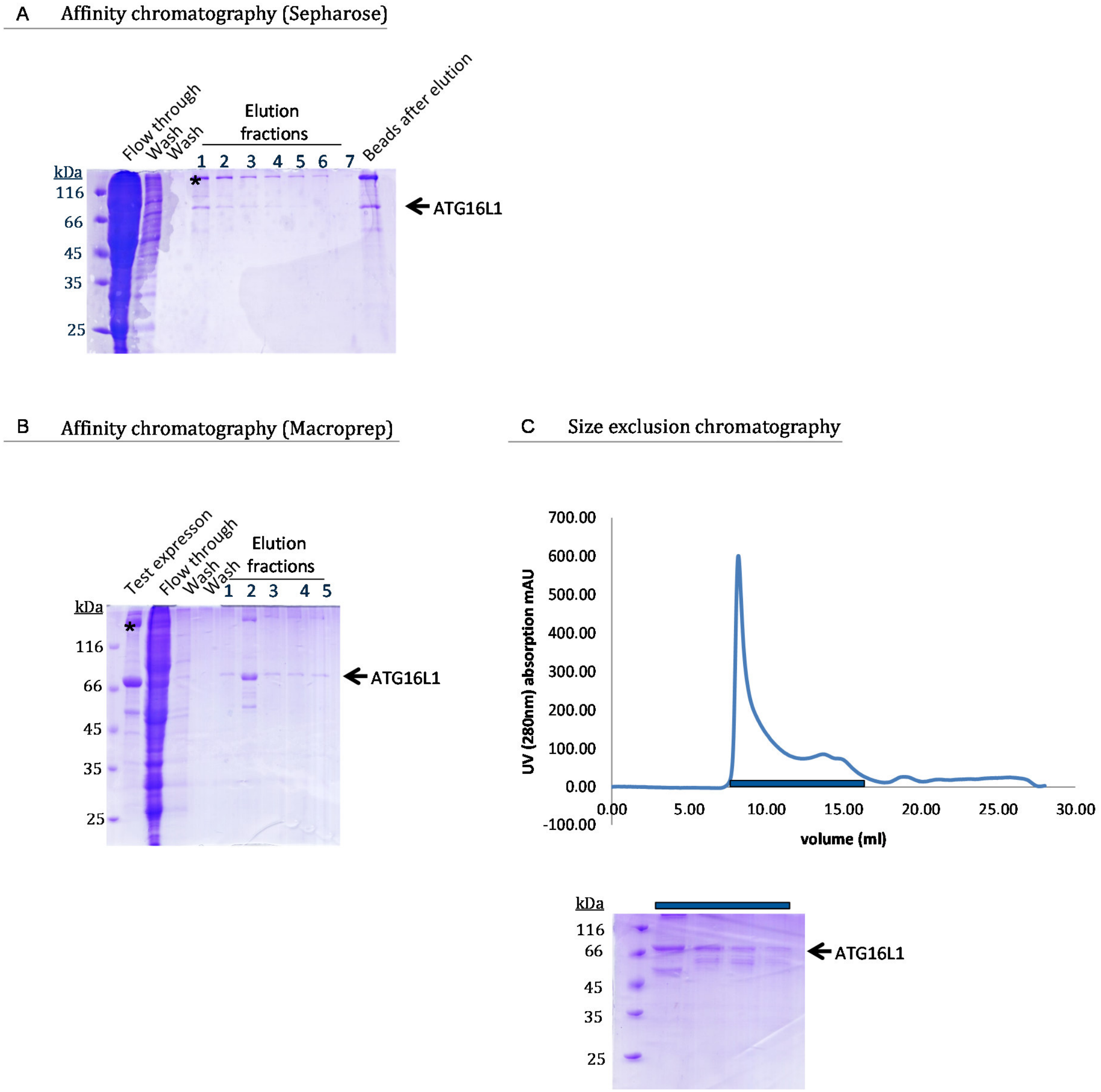


Figure 23: Large scale purification of Strep tagged full length human ATG16L1. A) 4 L HighFive cells ($1x10^6$ cells/ml) were infected with an aliquot of BIIC stock/1 L culture. Cells were harvested 3 days post infection. Cleared lysate was loaded on 2 ml Steptactin Sepharose gravity column, thoroughly washed and eluted with 5 mM desthiobiotin in 6x 5 ml factions. All samples were analyzed on 12 % SDS-PAGE B) 4 L HighFive cells ($1x10^6$ cells/ml) were infected with an aliquot of BIIC stock/1 L culture. Cells were harvested 3 days post infection. Cleared lysate was loaded on 2 ml Streptactin Macroprep gravity column, thoroughly washed and eluted with 5 mM desthiobiotin in 5x 5 ml factions. All samples were analyzed on 12 % SDS-PAGE. Second elution fraction was concentrated in Vivaspin 10000 MWCO to 0.5 ml (3 mg/ml) and loaded on S200 10/30 size exclusion chromatography column connected to Äkta Purifier system. Elution fractions were analyzed on 12 % SDS PAGE gel. Bands marked with asterisk (*) is an insect cell protein unspecific binding Strep beads.

Macroprep beads were shown most successful in the case of ATG16L1. Compared to HighCapacity and Sepharose, a fraction of the expressed protein was eluted from Macroprep beads. Upon loading of the post elution beads on the SDS PAGE gel, a band corresponding to the full length ATG16L1 could still be observed (Figure 23A, last lane and Figure 24A, first lane). Based on this, a part of the protein could be eluted with 5 mM desthiobiotin, but the majority of the protein still remained on the beads. Even upon incubation of the beads with 10 mM desthiobiotin and additional 5 elution fractions (Figure 24, lanes labelled 7-11), no additional protein was eluted. In an attempt to further enhance elution, half of the resin was incubated at 4°C overnight with 10 mM desthiobiotin. The other half was incubated with 100 µl TEV protease (1 mg/ml), at 4°C overnight. TEV protease would cleave off the tag, provided the cleavage site is accessible. This would allow the protein to detach from the beads. However, following the overnight incubation with either TEV or 10 mM desthiobiotin, protein did not elute from the beads (Figure 24B).

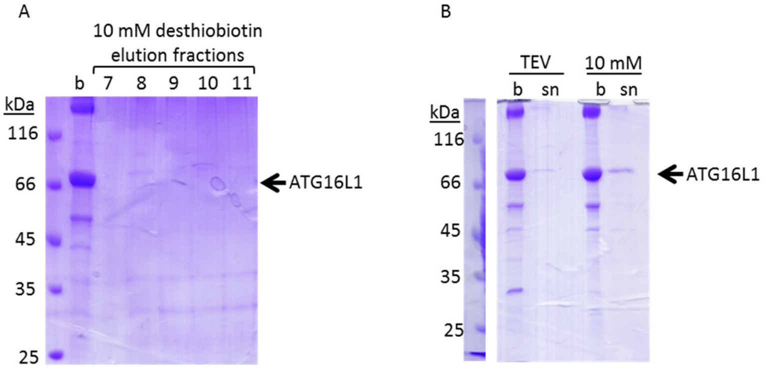


Figure 24: Additional elution attempts. A) Resin was eluted with 10 mM deshtiobiotin for another 5 fractions, before it was split. B) Half (1 ml) of the resin was incubated with 100 µl TEV, overnight at 4°C; the other half (1 ml) was incubated overnight with 10 mM desthiobiotin. Following the overnight incubation, resin (b) and supernatant (sn) were analyzed on 12 % SDS PAGE gel.

Strep purification is based on a strongest non-covalent bond known in nature, between avidin and biotin. Since avidin-biotin interaction cannot be broken, avidin is modified into Streptavidin, and a short peptide (Strep tag) is designed to bind in the biotin pocket. This interaction can be displaced by binding of desthiobiotin. Since human ATG16L1 was not eluted from the beads, it might be that the human ATG16L1 is biotinylated. Biotinylated protein would bind the beads directly, not via tag only. During elution with desthiobiotin, even if the Strep tag is displaced, the protein would remain on bound to the resin. To test for the biotinylation of human ATG16L1, a sample of full length ATG16L1 was subjected to a Western blot using AP conjugated avidin (Figure 25), which does not bind the Strep-tag. To narrow down the region of the ATG16L1 undergoing biotinylation, untagged C-terminal construct (ATG16L1_G303-Y607) was included as well. To exclude the possible effect of the Strep tag on the full length ATG16L1, another Strep tagged protein (Strep_ATG5~ATG12) was used as a control.

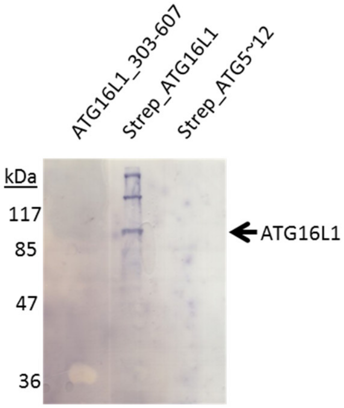


Figure 25: Western blot against biotin. Primary antibody: ImmunoPure Avidin, AP conjugated 1:500. Detection: AP colorimetric method using BCIP/NBT substrate.

Apart from the two high molecular weight bands (representing proteins from insect cells binding to Streptavidin, observed also in Figure 20) present in the purified Strep_ATG16L1 sample, a third band positive for biotinylation is ATG16L1. Since no biotinylation signal is seen for the construct 303-607, Western blot against biotin suggests the possible biotinylation occurs before position 303 of the human ATG16L1. To further pinpoint the biotinylation site, a sample was sent for mass spectrometry analysis. Mass spectrometry can detect post translational modification when they result in a size difference of peptides generated by tryptic digest. If the identification of the biotinylated site was possible, a mutant could be designed to circumvent this modification and hopefully enable elution of ATG16L1 from affinity resin. The SDS PAGE band from a fresh purification was sent to the CPRO Group for analysis and search for biotinylation sites. However, mass spectrometry studies did not yield any conclusive results.

Considering biotinylation site on human ATG16L1, purification of the Strep tagged ATG16L1 was attempted using Pierce® Monomeric Avidin Agarose (Thermo Fisher Scientific) resin, which enables purification of biotinylated molecules and shows no binding specificity towards the Strep tag. 4 L of HighFive cells were infected with BIIC stock of Strep tagged human ATG16L1. The cells were harvested 3 days post infection, and lysed according to the protocol described above. To ensure protein production, 10 ml of the culture were used for test expression with MacroPrep Strep beads.

The Avidin Agarose beads (1.5 ml) were primed with 1x PBS and 2 mM biotin in 1x PBS according to the manufacturer’s instructions. Biotin bound to reversible binding sites was washed with 0.1 M glycine, pH 2.8 and resin was equilibrated with 1x PBS. Beads were added to the supernatant containing full length ATG16L1 and incubated at 4°C, 1 hr, on a rocking platform. Unbound protein was washed with 1x PBS, and elution was done with 2 mM biotin, collecting 6x 1ml fractions. Additionally, beads were washed with 0.1 M glycine, pH 2.8, and 2x 1 ml fractions were collected. All samples were loaded on 12 % SDS PAGE for analysis (Figure 26).

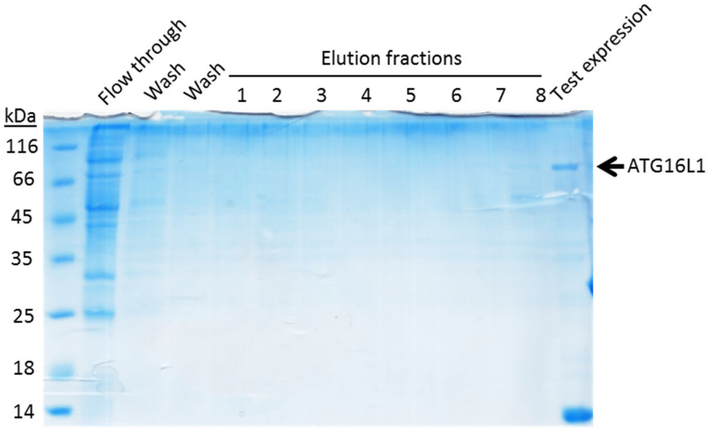


Figure 26: Avidin resin purification of Strep full length human ATG16L1. 4 L HighFive insect cells (0.8x10^6 cells/ml) was infected with BIIC stock. Cells were harvested 3 days post infection. Cleared lysate was loaded on 1.5 ml avidin resin. Purification was done according to the instructions provided by Thermo Fisher Scientific. Elution fractions 1-6 were eluted with blocking buffer (2 mM biotin in 1x PBS). Elution fractions 7 and 8 were eluted with 0.1 M glycine pH 2.8. Control: test expression using 10 ml of culture and Macroprep beads.

Purification of Strep tagged full length human ATG16L1 using Avidin resin resulted was not successful. The protein either did not bind to the beads, or was not eluted from the beads neither with 2mM biotin nor with 0.1M glycine, pH 2.8. However, based on test expression using 10ml of the culture and Macroprep beads, a band was observed on the beads corresponding to the size of full length human ATG16L1. This band was also identified as human ATG16L1 with tryptic digest and mass spectrometry (ATG16L1 with Score 135, next best hit: Rab GDP dissociation inhibitor with Score 42).

4.1.3.2 FLAG tagged full length human ATG16L1

Majority of the proteins are captured from the cellular content using affinity tags. Addition of tags to the protein sequence is a double edged sword. For some proteins, tags are important for solubility, and in some cases tags also participate in crystal contact formation. On the other hand, cleavage of solubility tags renders some proteins insoluble, thus unusable for structural studies. Along the same line, tags, no matter how small, can introduce enough disturbances in the process of crystal packing that some proteins crystallize only without them. Thus, the type of tag and its position has to be empirically tested to observe the effects it has on the protein or crystal. In this work, all fusion tags are separated from the protein sequence with TEV cleavage site (unless otherwise noted).

Recombinant expression of human ATG16L1 is not high enough to be able to separate the protein without an affinity tag. Since Step tag purification was not progressing, and His tag version resulted low yield and purity, full length ATG16L1 was cloned into a FLAG tagged vector. FLAG tagged ATG16L1 was expressed from insect cells, following the protocol described above. Cell lysate was incubated with FLAG beads for 3-4 hr, 4°C on a rocking platform. Purification protocol was performed according to the manufacturer's instructions.

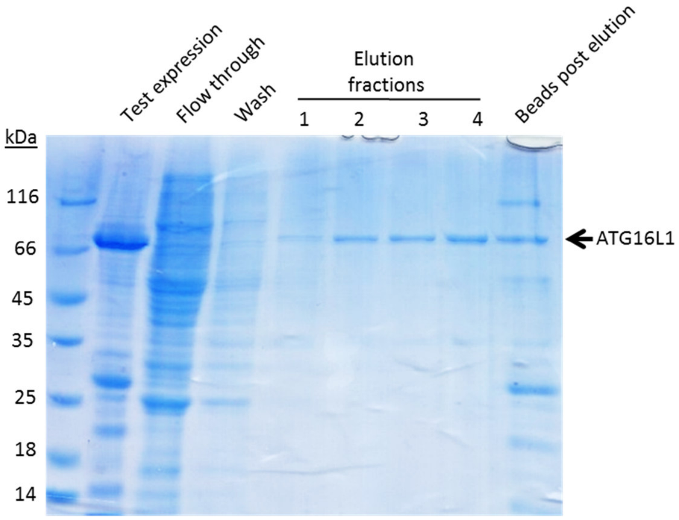


Figure 27: Purification of FLAG tagged full length human ATG16L1. 4 L of HighFive insect cell culture (0.8x10^6 cells/ml) were infected with P3 Flag full length human ATG16L1. Cells were harvested 3 days post infection. Cleared lysate was incubated with FLAG beads according to the protocol from BioLegend. Elution buffer contained 100 µg/ml FLAG peptide. All fractions were analyzed on 12 % SDS PAGE gel.

Elution fractions 2-4 were concentrated, however during the concentration process, all of the protein precipitated and could be collected by centrifugation at 14000 rpm at the bottom of the tube as a white pellet. FLAG beads were regenerated according to the manufacturer's protocol, and used for the fresh purification. It appeared the protein could not be removed from the beads completely and the next purification resulted in lower amount of protein eluting from the beads, and a greater amount of protein remaining on the beads, compared to the previous use. Even a more aggressive elution with 0.2 M glycine, pH 2.8 could not detach the protein.

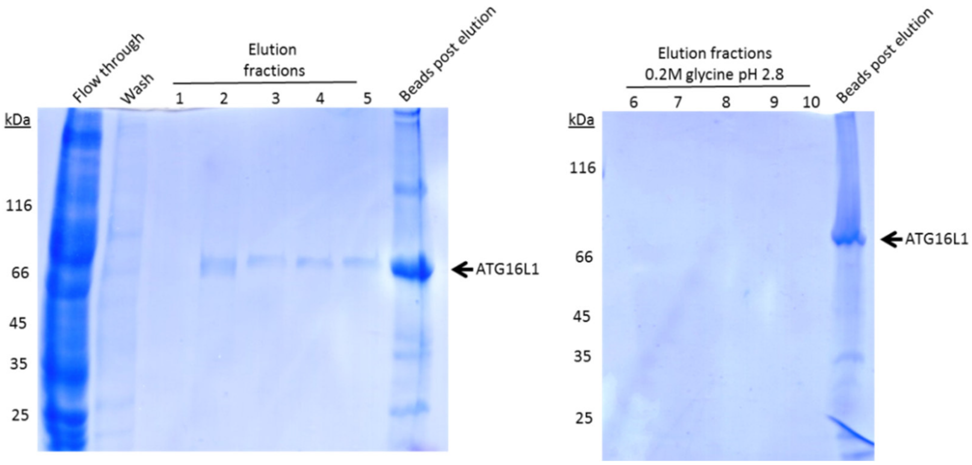


Figure 28: Purification of FLAG tagged full length human ATG16L1. 4 L of HighFive insect cell culture ($0.8x10^6$ cells/ml) were infected with P3 Flag full length human ATG16L1. Cells were harvested 3 days post infection. Cleared lysate was incubated with FLAG beads according to the protocol from BioLegend. Elution buffer containing 100 µg/ml FLAG peptide was used for fractions 1-5. Fractions 6-10 were eluted with 0.2 M glycine pH 2.8 and immediately neutralized with 1 M Tris pH 8. All fractions were analyzed on 12 % SDS PAGE gel.

Since stability and purity of the sample are important determinants for successful protein crystallization, the small amount of ATG16L1 eluted from FLAG beads was concentrated to 1 mg/ml and used for biophysical studies in hope to understand the nature of the protein better and find a way to obtain enough pure, soluble and homogenous protein for crystallization.

4.1.4 Biophysical studies of full length human ATG16L1

4.1.4.1 ThermoFluor of full length human ATG16L1

Protein solubility is influenced by many factors among which pH, ionic strength and presence of other components play a major role. At a pH close to the isoelectric point (pI), protein usually has the least solubility. Calculated pI of the full length ATG16L1 is 6.2. Buffers used for purification steps (both lysis and elution) were above pH of 6.2. At the pH above the pI, the net charge of the protein is negative, and more water molecules can interact with it, keeping it soluble. In the case of ionic strength, salt ions interact with oppositely charged groups on the protein, causing more interactions between a protein and the solvent than between protein molecules themselves. However, screening the effects different pH values and ionic strengths have on protein solubility can be of great advantage.

ThermoFluor assay, a fluorescence-based thermal stability method, monitors the effects of different buffers on the stability of the protein. A hydrophobic fluorescent dye is mixed with the protein solution and upon thermal denaturation the dye binds to the hydrophobic patches resulting in an increased fluorescence. The fluorescent signal is monitored as a function of temperature.

Due to the low stability and solubility in the elution buffer during purifications, FLAG tagged human full length ATG16L1 was subjected to the ThermoFluor buffer test to identify buffer conditions that increase the stability of the protein. Low amount of available pure protein did not allow for ThermoFluor conditions to be optimized. Nonetheless, the ThermoFluor assay showed two melting curves, for 50 mM HEPES pH 7, 100 mM NaCl and for 50 mM HEPES pH7, 200 mM NaCl, both with melting temperature of 54°C.

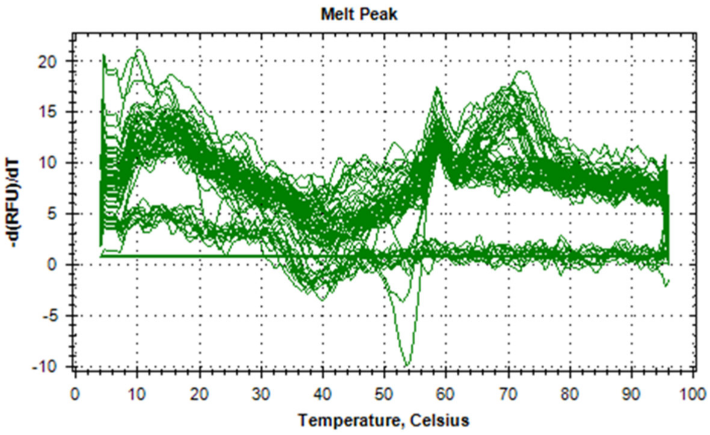


Figure 29: ThermoFlour melting curves for FLAG full length human ATG16L1. 1 mg/ml of protein was mixed with 10x SyproOrange and aliquoted into ThermoFluor Buffer screen.

4.1.4.2 Circular dichroism (CD) spectroscopy of full length human ATG16L1

The small amount of purified FLAG tagged full length human ATG16L1 was enough for recording a circular dichroism (CD) spectrum. The spectrum was recorded from 190 nm to 260 nm at 20°C.

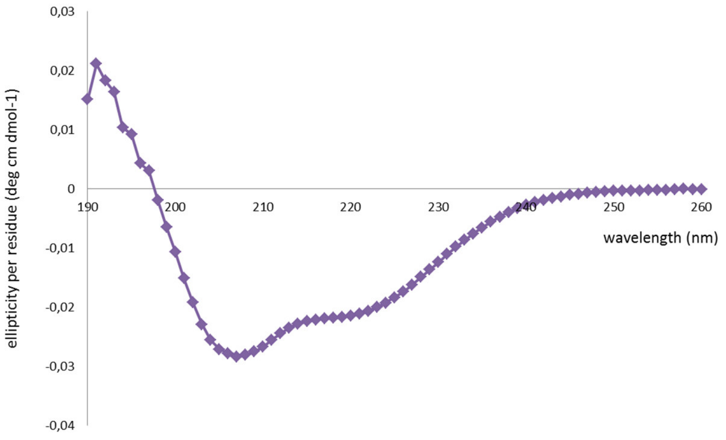
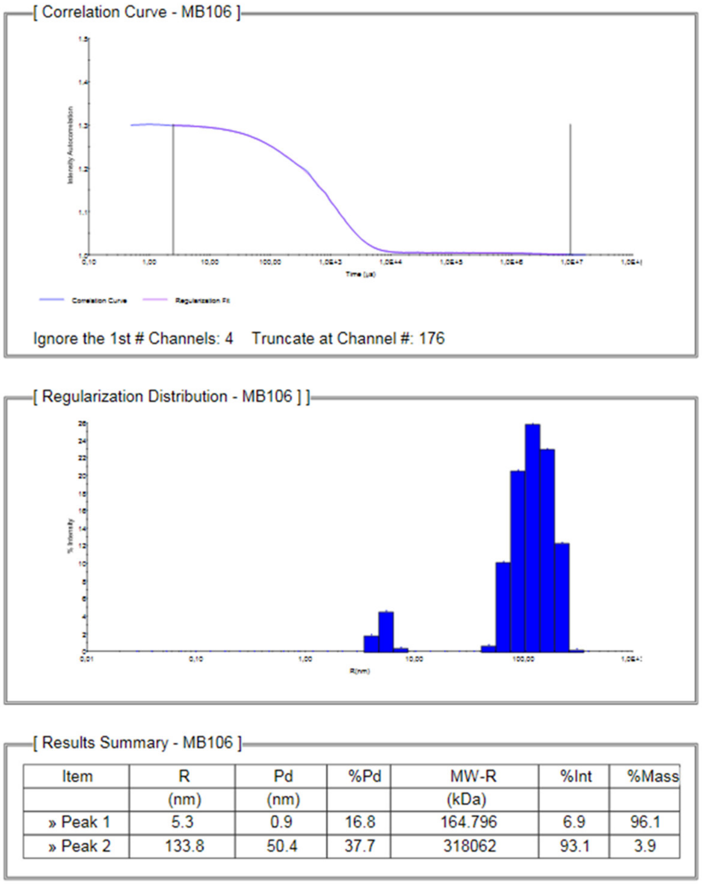


Figure 30: Circular dichroism curve for full length ATG16L1. 300 µl of 1 mg/ml of protein was dialyzed overnight in 2 L CD buffer. 0.8 mg/ml of protein was recovered in 350 µl and a CD spectrum was measured from 190 nm to 260 nm at 20°C.

Based on CD spectra of the full length human ATG16L1, the protein appears to be folded. A maximum at 190 nm together with the two minima at 208 nm and 222 nm are a typical features of α-helical secondary structure. K2D2 (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2//) calculated 11.53 % of the protein are α-helical and 28.35 % are β-strands. However, these values have to be considered with caution. CD spectra reflect an average over the entire molecule, and the percentage of secondary structure elements is not an absolute assessment. What the CD spectra of the full length human ATG16L1 does show with certainty is that the protein is folded with distinguishable secondary structure elements.

4.1.4.3 Dynamic light scattering (DLS) of full length human ATG16L1

Same sample used for CD was subsequently used for DLS. Every macromolecule has an ability to scatter light and this scattering is dependent on the overall size of the particle. Size distribution of the particle in solution as well as homogeneity of the sample can be studied using DLS. Homogeneity is crucial for crystallization of a protein, since a sample with mixed population will hardly arrange itself into well-ordered periodic assemblies such as a crystal. Based on the scattering of the laser light, full length human ATG16L1 has two populations. Large portion of the protein in solution (96.1 %) has a molecular weight of 165 kDa, which is close to a molecular weight of a dimer (monomer is around 70 kDa in size). This makes sense because the coiled-coil region of the ATG16L1 is responsible for dimerization of the protein. A small portion of the sample (3.9 %) is a large aggregate or oligomer with an average molecular weight of 318 000 kDa.



Ignore the 1st # Channels: 4   Truncate at Channel #: 176

Item	R	Pd	%Pd	MW-R	%Int	%Mass
	(nm)	(nm)		(kDa)		
» Peak 1	5.3	0.9	16.8	164.796	6.9	96.1
» Peak 2	133.8	50.4	37.7	318062	93.1	3.9

Figure 31: Dynamic light scattering analysis full length ATG16L1

The sample used for both CD and DLS was shock frozen in liquid nitrogen and stored at -80°C. Eventually, the sample was thawed, centrifuged at 10 000 rpm, 15 min, 4°C and sent for SAXS analysis at ESRF. However, the measured scattering curve was uninterpretable.

4.1.4.4 Co-expression of ATG16L1 with ATG5/ATG12 conjugate

Difficulties with purification of full length ATG16L1 might be attributed to its tendency towards aggregation. Full length ATG16L1 could not be successfully eluted from affinity columns, with only a small fraction of the protein in the elution, and majority of the protein remaining on the resin. This could be explained by formation of large insoluble aggregates. If the protein precipitates on the affinity resin, it will not be eluted with standard methods. However, even the small part of soluble protein which was eluted from the affinity resin was in the void volume of size exclusion, indicating formation of large aggregates. As mentioned above, protein solubility depends on pH, ionic strength and other components present in solution. The effect of pH and salt ions can be tested by a ThermoFlour assay. The category of other components often include organic solvents (PEGs or glycerol) or reducing reagents (such as BME or DTT) as well as known interaction partners. Interaction partners often solubilize proteins by masking the hydrophobic, solvent exposed regions, which, in the absence of an appropriate binding partner, lead to aggregation of the protein. To test this, Strep tagged full length ATG16L1 was co-expressed with its well-known partners ATG5~12. Expression and purification of ATG5~12 has been optimized by Archna in our lab. Strep tagged ATG16L1 with His-ATG5/GST-ATG12 and untagged ATG7/ATG10 (ATG7/ATG10 are required for ATG5~12 conjugation) were co-expressed together in insect cells. Since insect cells have their own GST protein, first purification step was Ni-NTA to avoid oversaturating GST beads with endogenous protein. Elution fraction from Ni-NTA was subsequently incubated with GST beads (Figure 32 A and B). To confirm the presence of ATG16L1, GST elution fraction was subjected to Gene TAGnology Strep-tag AP Chromagenic Detection kit (iba BioTAGnology) (Figure 32C).

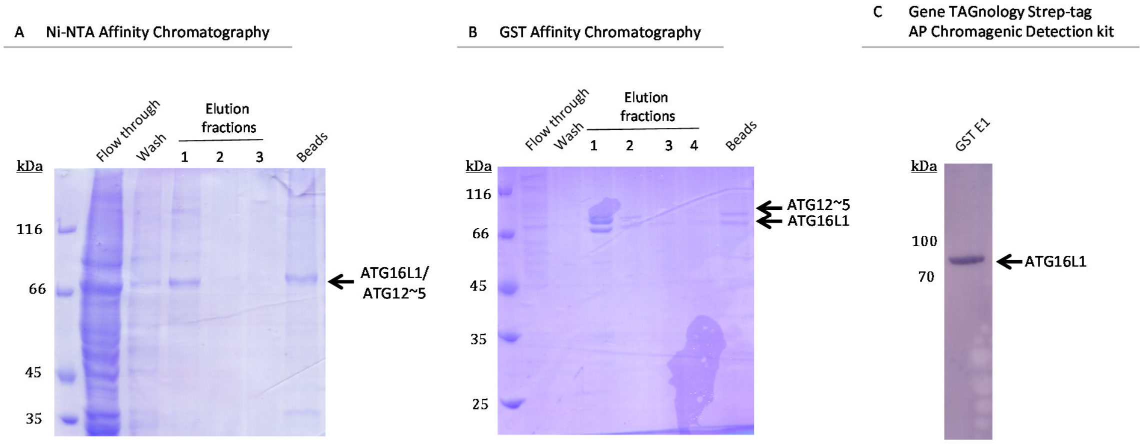


Figure 32: Co-expression of ATG16L1 with ATG12~5. 1 L of HighFive insect cell ($0.8x10^6$ cells/ml) was infected with three viruses: one containing His-ATG5/GST-ATG12, one with untagged ATG7 and ATG10 and BIIC stock of Strep-ATG16L1. Cells were harvested 3 days post infection. A) First affinity step was Ni-NTA: protein was eluted with 100 mM Imidazole. Beads after elution were loaded as control of elution efficiency. B) First elution fraction from Ni-NTA affinity was incubated with 2 ml GST beads. GST elution was performed with 20 mM reduced glutathione in 4x 3ml fraction. Beads after elution were loaded as control of elution efficiency. C) GST elution fraction 1 (E1) was loaded on a separate gel and Strep tag on ATG16L1 was detected using Gene TAGnology Strep-tag AP Chromagenic Detection kit.

Thus, when co-expressed with ATG12~ATG5, ATG16L1 expression as well as its elution was also successful. When the purification principle was shown to work, the culture volume was increased to 6 L and repeated. Fresh virus for Strep tagged ATG16L1 was produced for infection to eliminate possible delay in virus release by BIIC. Elution fraction after GST purification was concentrated and loaded on Superose 6 10/30 column connected to the Äkta Purifier system. Fractions were analyzed on 12 % SDS PAGE gel (Figure 33). Most of the protein was present in the void volume. Thus, problem of ATG16L1 aggregation was not solved by co-expression with its interaction partners.

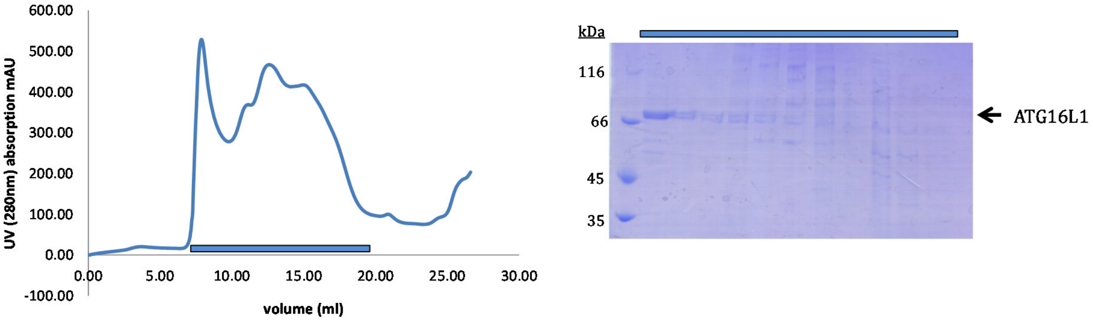


Figure 33: Co-expression of Strep ATG16L1 with His-ATG5/GST-ATG12. 3 mg/ml of the complex following GST affinity were loaded on Superose6 10/30 gel filtration column. All fractions were analyzed on 12 % SDS PAGE gel.

4.1.5 Phosphomimetic variants of ATG16L1

Apart from the classical role of ATG16L1 in complex with ATG5~ATG12 in autophagosome formation, this protein has additional roles, not necessarily always exclusive to autophagy. A mutation from threonine to alanine on position 300 has been linked to Crohn's disease. This variant was tested to see if it has an effect on the interactions with DDB1. Since threonine is one of the classical residues to undergo phosphorylation, phosphomimetic substitutions T300E and T300D were introduced. Glutamate (E) and aspartate (D) are both negatively charged amino acids and mimic a phosphorylated site. In addition, four more positions on the human ATG16L1 were predicted to be phosphorylated: S287, S289, S290, and S304. Each of these is substituted with a glutamate. Viruses harboring S287 and S290 could be successfully amplified to P2 level. Interaction of the phosphomimetic variants (T300E, T300D, S287E and S304E) with DDB1 was tested by co-infection and pull down assay (Figure 34). In the pull down assay, tested phosphomimetic variants of ATG16L1 behave similarly to the wild type, in that they show no interaction with DDB1.

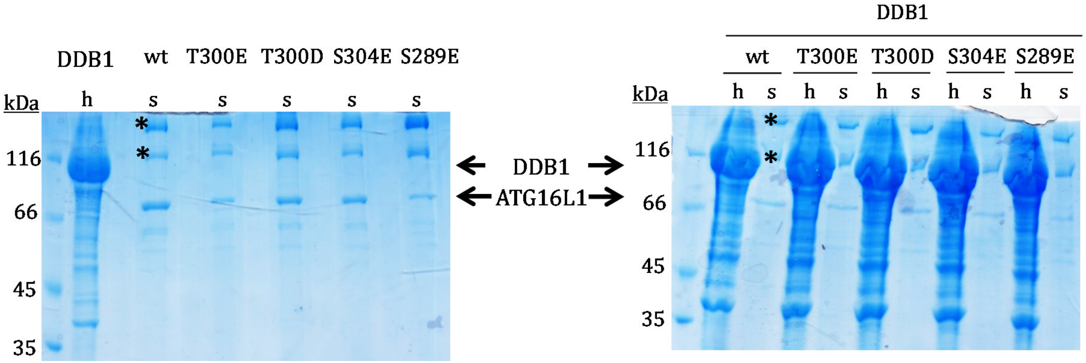


Figure 34: Co-expression of His-DDB1 with Strep-ATG16L1 constructs (wt: ATG16L1_M1-Y607, T300E: ATG16L1_M1_T300E_Y607, T300D: ATG16L1_M1_T300D_Y607, S304E: ATG16L1_M1_S304E_Y607, S287E: ATG16L1_M1_S289E_Y607) in 20ml Sf-21 (1x10^6 cells/ml). Cells were harvested 3 days post infection. Samples were analyzed on 12% SDS-PAGE gels (h= His pull down; s= Strep pull down). Bands marked with asterisk (*) is an insect cell protein unspecific binding Strep beads.

This concludes the attempts towards expression and purification of the full length human ATG16L1. Despite reasonable expression of the protein in small scale from insect cells, high yield and purity of full length human ATG16L1 which is needed for crystallographic studies and further biochemical characterization could not be obtained during the time period of this work.

4.1.6 Expression and purification of C-terminal domain of human ATG16L1

Constructs containing C-terminal domain of human ATG16L1 were designed as described in Section 4.1.1.1.2 and cloned with either His or Strep tag into both bacterial and insect cell expression vectors. Test expression from *E. coli* was previously done by Petra Völler in our lab, with no success in obtaining soluble expressed protein. Baculoviruses containing C-terminal constructs were tested in insect cells. SDS PAGE gel displaying P2 test expression of C-terminal ATG16L1 constructs from insect cells is shown in Figure 35.

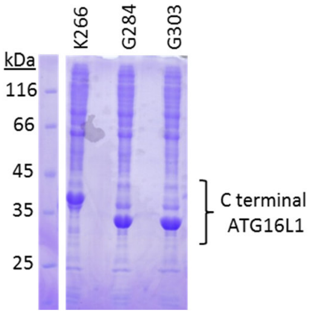


*Figure 35: SDS PAGE of small scale test expression of His tagged C-terminal constructs (**K266**-Y607, **G284**-Y607 and **G303**-Y607) of human ATG16L1 produced with P2 virus in Sf-21 cells.*

C-terminal constructs of human ATG16L1 were successfully expressed with a His tag and P3 virus was used for large scale production in HighFive cells. Purification protocol consists of two steps: affinity chromatography and gel filtration. Since affinity purification for His tag is not very specific, impurities were removed by gel filtration, yielding a highly soluble and pure C-terminal ATG16L1 construct. Gel filtration retention volume indicates a monomeric protein in solution. All of the C-terminal constructs were purified using the same procedure. Sample purification for His tagged ATG16L1_G284-Y607 is shown in Figure 36.

In the case of ATG16L1, the single domain shows higher solubility and stability compared to the full length multi-domain protein. In this work, main focus is given to C-terminal domain of human ATG16L1. Following gel filtration step, C-terminal constructs of human ATG16L1 were concentrated to 25 mg/ml and either frozen in liquid nitrogen and stored at -80°C, or used for crystallization screens. Trials were set with 5 and 10 mg/ml of protein with and without a tag, at both 4°C and 20°C. The initial screens did not result any protein crystals.

First major bottleneck in structural biology was already mentioned during construct design and cloning, hoping for a soluble and homogeneous construct with high production yield. Second major bottleneck is obtaining crystals of that protein. Since initial crystallization trials with C-terminal constructs of ATG16L1 were not successful, protein is further analyzed using biochemical and biophysical methods to assess the quality of the protein and improve its crystallization.

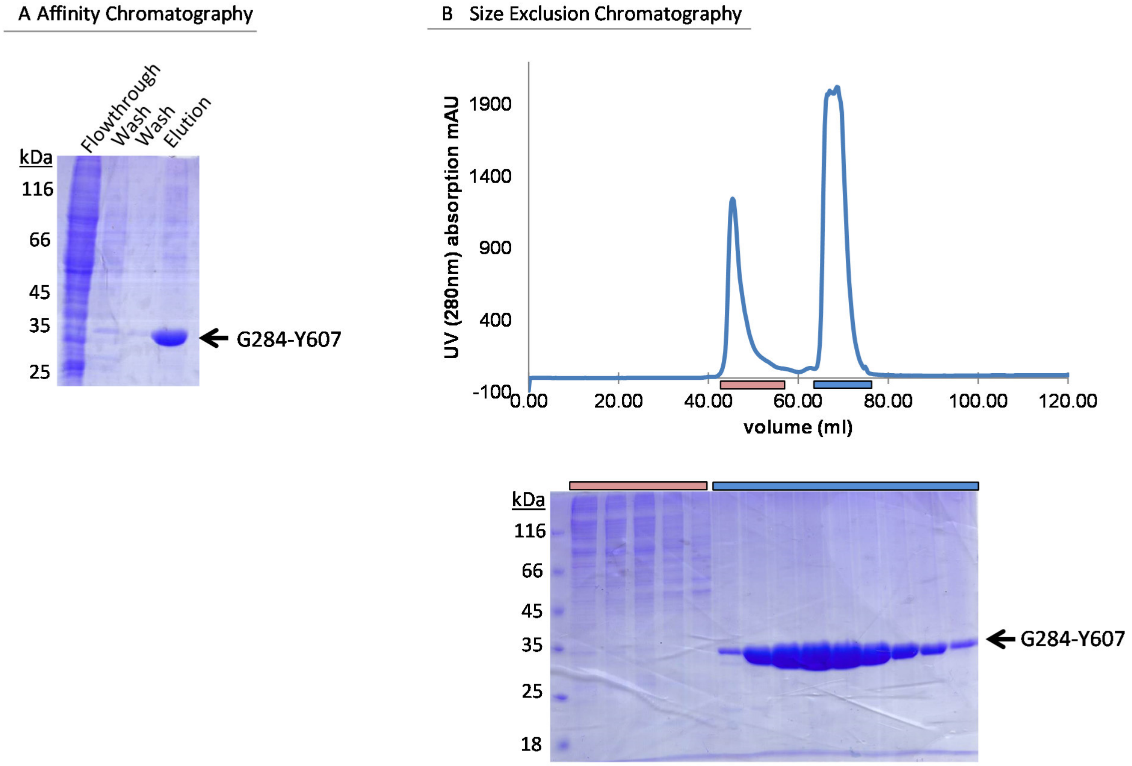


Figure 36: Purification procedure for C-terminal constructs of human ATG16L1. 1 L culture of High Five insect cells (0.8x10^6 cells/ml) was infected with 5-10 % P3 virus coding for His tagged C-terminal domain of human ATG16L1. Affinity chromatography was done with Ni-NTA beads; bound protein was eluted with 200 mM Imidazole. Eluted fractions were concentrated and loaded on Superose 75 16/60 connected to the Äkta Purifier system. Fractions were analyzed on SDS PAGE gel, pooled, concentrated and shock frozen in liquid nitrogen or used for crystallization trials. Purification for His_ATG16L1_G284-Y607 is represented here.

4.1.7 Biophysical studies of C-terminal domain of human ATG16L1

4.1.7.1 ThermoFluor assay with C-terminal domain of human ATG16L1

C-terminal domain of human ATG16L1 was purified as described and subjected to the ThermoFluor assay, to study effects pH and salt concentration have on the protein.

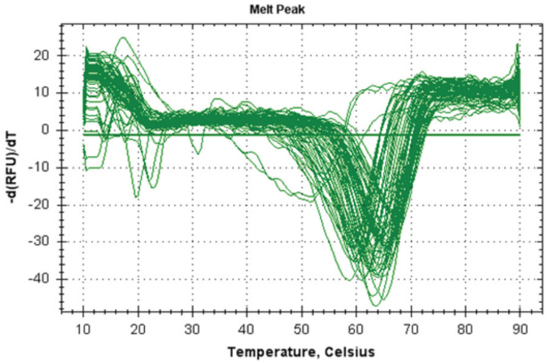


Figure 37: ThermoFluor melting curves for C-terminal domain of human ATG16L1. 2 mg/ml of protein was mixed with 10x SyproOrange and aliquoted into Thermofluor Buffer screen.

According to the melting curves recorded (Figure 37), it was difficult to determine one single best condition, since many buffers gave similar melting curves resulting in the same level of protein stability. Buffers in the pH range between 6 and 8, with NaCl concentration between 200 mM and 500 mM keep the melting temperature of the protein between 62°C and 66°C. The buffer used for purification of C-terminal domain of human ATG16L1 falls within the range observed as optimal by ThermoFluor.

4.1.7.2 Circular dichroism (CD) and dynamic light scattering (DLS) with C-terminal domain of human ATG16L1

Secondary structure composition and folding properties of the C-terminal constructs of ATG16L1 were investigated using circular dichroism. The spectrum was recorded from 190 nm to 260 nm at 20°C.

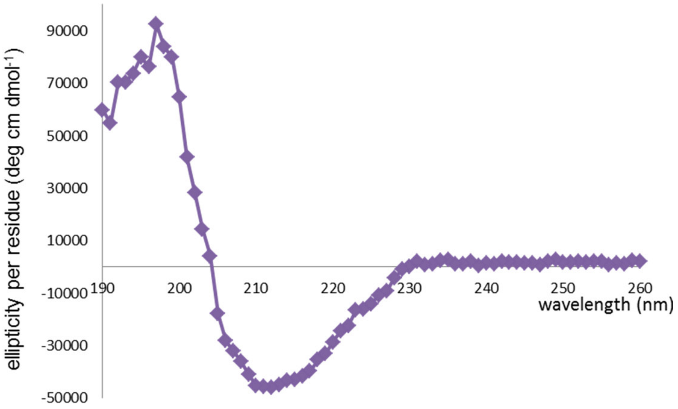
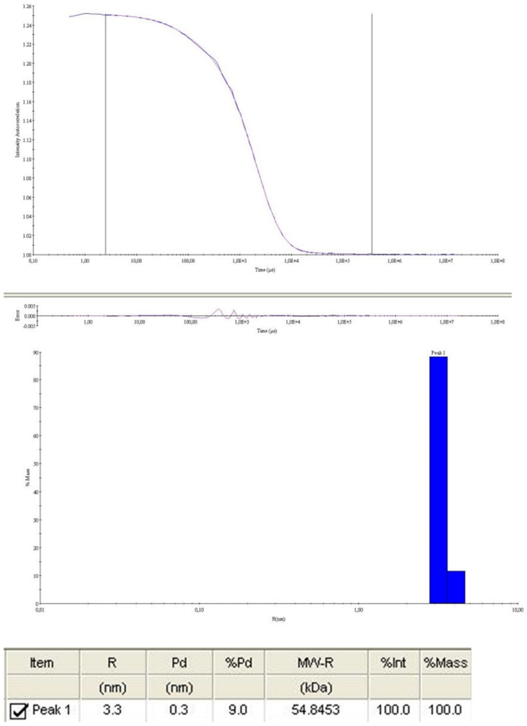


Figure 38: Circular dichroism curve for C-terminal domain of human ATG16L1 (G303-Y607). 300 µl of 2 mg/ml of protein was dialyzed overnight in 2 L CD buffer. 1.6 mg/ml of protein was recovered in 350 µl and a CD spectrum was recorded from 190 nm to 260 nm at 20°C.

CD spectra reflect an average over the entire molecules. Based on the curve for C-terminal domain of ATG16L1, protein was properly folded (Figure 38). The overall shape of the recorded spectra represents a typical β-sheet composition. A typical β-sheet has a characteristic minimum at around 217 nm and a maximum around 198 nm. This confirms the prediction that secondary structure of C-terminal domain of human ATG16L1 is mainly β-sheets. K2D2 (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2//) calculated 12.53 % of the protein is α-helical and 35.2 % are β-strands.



Item	R	Pd	%Pd	MW-R	%Int	%Mass
	(nm)	(nm)		(kDa)		
☑ Peak 1	3.3	0.3	9.0	54.8453	100.0	100.0

Figure 39: Dynamic light scattering analysis of C-terminal domain of human ATG16L1 (G303-Y607)

Apart from good solubility upon expression and purification, protein must also be homogeneous, ideally as a single population, for crystallization. Dynamic light scattering was used to determine size distribution of protein sample in solution. According to the DLS (Figure 39), C-terminal domain of human ATG16L1 exists as a monomer in solution, which is in agreement with the gel elution profile of the protein.

4.1.8 In search of interaction partners of C-terminal domain of human ATG16L1

As already mentioned in the case of full length ATG16L1, co-expression of the target protein with its interaction partners might boost production levels and solubility. In case of full length ATG16L1, interaction partners were used in an attempt to solubilize the protein and possibly enhance its expression. C-terminal domain of human ATG16L1 was expressed in high yield and was quite soluble. However, crystal growth was not very promising. Following the same logic as for co-expression, co-crystallization would provide new surfaces for crystal contacts and might lead to crystal packing. This was the hope for the C-terminal domain of human ATG16L1. Not only would the interaction partners help in crystallization, but a crystal structure of the complex would provide molecular detail of the mode of action of C-terminal part of human ATG16L1 and help define the function of this domain more precisely.

C-terminal half of the human ATG16L1 is present only in higher eukaryotes and is not essential for canonical autophagy process. Literature search suggested a few candidates for interaction partners of specifically this domain. We have chosen to focus our attention on the following list: DDB1, CARD domain of NOD1, CARD1 domain of NOD2, TIR domain of TLR2, cytosolic domain of TMEM59.

4.1.8.1 Pull-down assays

Since most of the interaction partners reported in the literature for C-terminal domain of human ATG16L1 are identified using immunoprecipitation method, one way to confirm those results is by a pull-down assay. Pull-down assay is based on affinity principle, and is used to either enrich a population of a certain protein or to detect interaction partners. In the case when interaction partners are known, both the target protein and the partner are tagged with different tags. Proteins are captured through the affinity tag on the affinity resin specific for the tag. Immobilized proteins are incubated with the source of interaction partner, either cell lysate or solution with purified protein. Affinity resin is washed and at the end loaded on the SDS PAGE gel for analysis.

Apart from DDB1, putative interaction partners of C-terminal domain of human ATG16L1 are relatively small domains, and could be expressed from *E. coli*, and will be presented first. The longest construct (K266-Y607) of the C-terminal domain of ATG16L1 with a Strep tag was purified from insect cells are described above. Purified Strep_ATG16L1-K266-Y607 was mixed with the lysate of *E. coli* culture expressing the respective interaction partners: His-TLR2, GST-NOD2 and GST-TMEM59. For the putative interaction partners expressed as GST-fusion constructs, GST alone was used as control to analyze unspecific binding via the large GST-tag. Following a 1hr incubation at 4°C, the mixture was split and one half was incubated with Strep beads, and the other half with the beads corresponding to the tag of the interaction partner. After another 1hr incubation at 4°C, beads were washed and loaded on SDS PAGE gel for analysis (Figure 40).

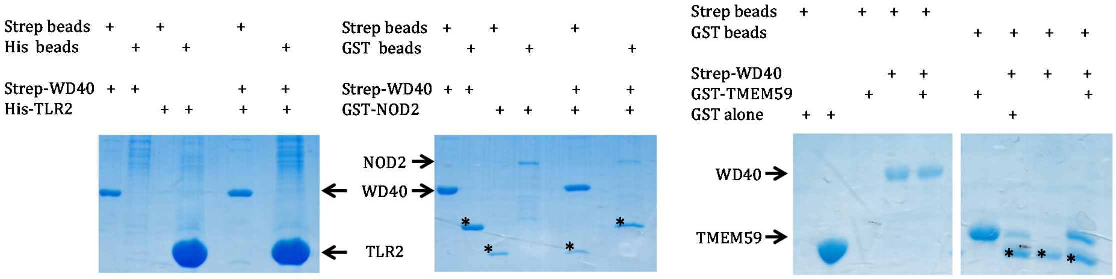


Figure 40: Pull-down studies with putative interaction partners of C-terminal domain of human ATG16L1. Strep G284-Y607 was purified from HighFive cells. Interaction partners (His-TLR2 (626-784), GST-NOD2 (1-209) and GST-TMEM59 (263-323)) were expressed from 50 ml E. coli, each according to their individual optimal expression conditions. Pellets from each expression were lysed and mixed with beads according to the figure above. Pull down samples were analyzed on 15 % SDS-PAGE gel. Bands marked with asterisk (*) are contamination/unspecific protein bands.

The previously reported and published interaction of ATG16L1 WD40 with TIR domain of TLR2, CARD domains of NOD2 and cytosolic domain of TMEM59 could not be confirmed using the pull-down assay. This could be attributed to the relatively low sensitivity of the pull-down method. Usually, stable protein:protein complexes with high affinity can be identified using pull-down assay. However, complexes with low affinity or transient interactions are more challenging to detect and often broken during washing steps in pull-down assay.

To test interaction between DDB1 with C-terminal domain ATG16L1, I took advantage of the fact that their expression system is the same. His tagged DDB1 was co-expressed with all constructs of ATG16L1 (three C-terminal constructs and full length protein) and their interaction was tested by pull – down assay. As a control, DDB2 co-expression was included, a DCAF protein which is known to form a stable DDB1-DDB2 complex. On the left hand side in Figure 41 are single expressions of all proteins, and on the right hand side is the gel with co-expressions.

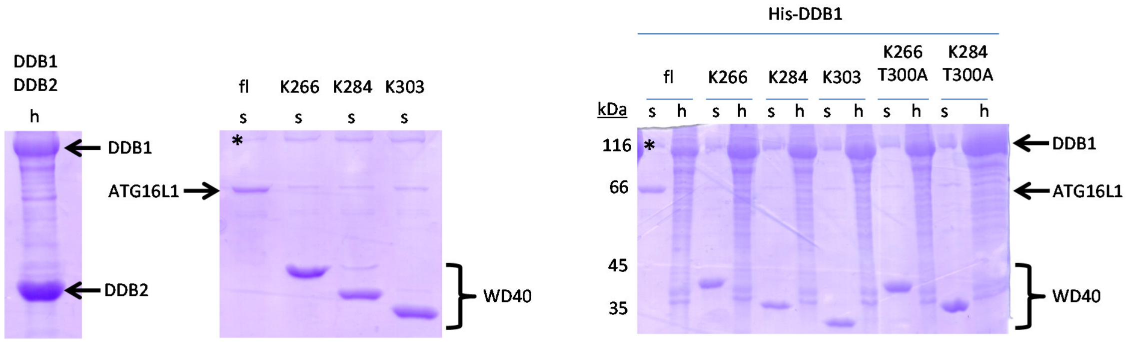


Figure 41: Co-expression of His-DDB1 with Strep-ATG16L1 constructs (fl: full length, K266: K266-Y607; K284: K284-Y607; K303: K303-Y607) in 20 ml Sf-21 (1x10^6 cells/ml). Cells were harvested 3 days post infection. Samples were analyzed on 12 % SDS-PAGE gels. Control: His-DDB1 and His-DDB2. Bands marked with asterisk (*) are proteins from insect cells which unspecifically bind to Strep beads. Legend: h = Ni-NTA beads, S = Strep-Tactin Macroprep beads.

Based on interaction studies between DDB1 and ATG16L1 using pull-down assay, complex formation between these two proteins could not be detected. In addition to low sensitivity, pull-down assay also require one or both proteins to be immobilized, which might interfere with complex formation. Size exclusion chromatography allows protein to be free in solution. Change in protein size due to complex formation is detected by the shift of the elution peak. Single protein will elute at one retention volume, while the complex will elute earlier, indicating a larger molecular weight. The shift between a large protein alone and a large protein bound to a small domain will not be as prominent as a shift between a small domain alone and a small domain bound to a large protein.

With this in mind, C-terminal domain of human ATG16L1 starting at position G284 was expressed from insect cells. Following affinity chromatography, protein is concentrated and loaded on analytical S200 10/30 gel filtration column. This run is shown in dark blue in Figure 42. To test the interaction with DDB1, equal molar amounts of the purified C-terminal domain are mixed with DDB1. Additionally, 2% TEV (incubated with the sample at 4°C, 6 hr) was added to cleave off tags, in case they interfere with the complex formation. To keep variation of external factors at the minimum, both runs were performed on the same purification system, on the same column, using the same running buffer, and one right after the other. Elution fractions were analyzed on SDS PAGE gel.

Elution profiles were overlaid to observe if there is any shift of the retention volume of the C-terminal domain of ATG16L1 when allowed to interact with DDB1. However, the elution volume of ATG16L1_G284-Y607 is the same with and without DDB1. Additionally, in case of interaction, bands for both proteins should be present in the same elution fraction, in the same amounts, when tested on the SDS PAGE gel. However, protein band corresponding the DDB1 is present only in the early peak, while ATG16L1_G284-Y607 is in the second peak, indicating no co-elution of the proteins on size exclusion chromatography.

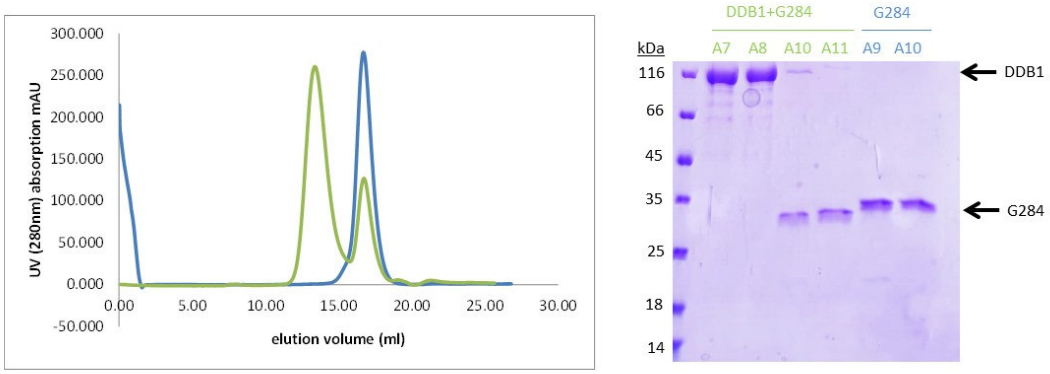
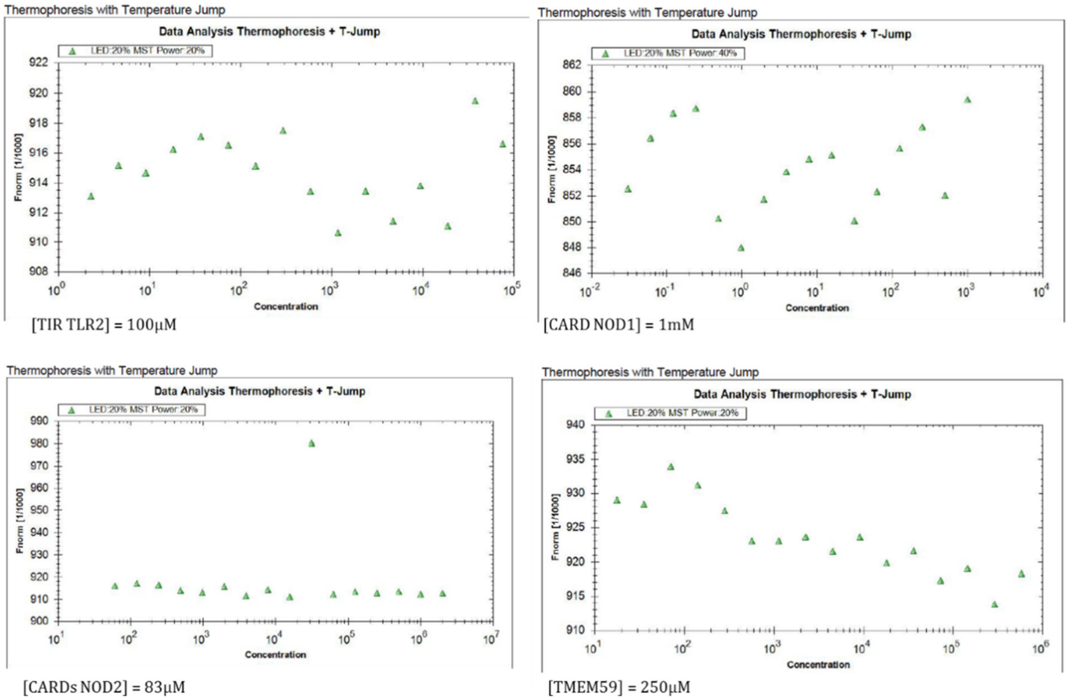


Figure 42: Analytical gel filtration of DDB1 and C-terminal human ATG16L1_G284-Y607 (1 mM of DDB1 and 1 mM of C-terminal human ATG16L1), green curve. Control: C-terminal human ATG16L1 alone (2.5mM), blue curve. Elution profiles are overlaid and fractions were analyzed on 12 % SDS-PAGE gel.

4.1.8.2 Microscale thermophoresis

Pull-down assay and size exclusion chromatography (SEC) are able to detect only strong interactions between proteins. Weaker interaction could be broken and disrupted during washing steps or on the column. Since there is no evidence to the affinity between C-terminal domain of ATG16L1 with any of the mentioned interaction partners, pull-down assay and SEC might not be a suitable method of detection in this case. A more delicate way to detect protein interaction is by microscale thermophoresis (MST). MST requires relatively low amount of protein and allows proteins to be free in solution during measurement. During an experiment, a microscopic temperature gradient is applied using infrared laser. Migration of molecules along the temperature gradient is measured and quantified. A change in protein migration rate along the temperature gradient is caused by an increase in size and change in charge or hydration shell during complex formation. C-terminal domain of ATG16L1 (G284-Y607) was labeled according to the NanoTemper instructions, and the interaction with proposed, unlabeled binding partners (His-TLR2, GST-NOD2, NOD1-His and GST-TMEM59) was tested using MST.

MST was repeated varying buffer conditions: salt concentrations and pH were varied to test an effect on the interaction. However, graphs recorded in different buffers show the same trend as the graphs in Figure 43. Based on these results, C-terminal domain of ATG16L1 does not form a direct interaction with proteins tested here.



[TIR TLR2] = 100µM

[CARD NOD1] = 1mM

[CARDs NOD2] = 83µM

[TMEM59] = 250µM

Figure 43: Thermophoresis experiment of C-terminal domain of human ATG16L1 with putative interaction partners (His-TLR2 (626-784), GST-NOD2 (1-209), NOD1-His (1-106) and GST-TMEM59 (263-323)). C-terminal domain of human ATG16L1 (G284-Y607) was labelled according to the NanoTemper protocol at concentration of 20 µM. Each point on the graph represents 1:1 dilution of the previous concentration. Maximum (starting) concentrations of each interaction partner are indicated below the corresponding thermophoresis graph.

So far interactions between C-terminal domain of ATG16L1 and its putative partners were tested in vitro, in buffer systems and in cellular systems such as insect cells. In case of strong, stable interactions, this level of testing would be enough to detect the complex formation and use it in structural studies. However, this was not the case, indicating a possibility of another element essential for the interactions with C-terminal domain of ATG16L1 but missing in the experimental systems used so far. To search for additional partners in a more comprehensive system, transient transfection of HEK293 6E was considered and tested. Due to time limitations and complexity of the autophagy process, no conclusive results can be reported at this time.

This concludes attempts to confirm known and putative interaction partners of ATG16L1. To achieve the aim of this project, and crystallize the C-terminal domain of ATG16L1, additional approaches were considered. Since the search for interaction partners was unrewarding, further engineering strategies at the protein level, such as limited proteolysis and surface entropy reduction mutagenesis, were instigated.

4.1.8.3 Limited proteolysis

One method for protein engineering and modification of already expressed, soluble protein is by limited proteolysis. Since the construct boundaries used in this study are based on prediction methods, it is possible that flexible regions are still present which might prevent crystal formation. Limited proteolysis provides information about the compactness and proteolytic stability of C-terminal domain of human ATG16L1. Each construct was purified as described above, and subjected to 12 proteases from the Proti-Ace Kits 1 and 2 from Hampton Research. The hope of this approach was that the small amount of proteases would cleave off flexible loops and ends, yielding a more compact fragment.

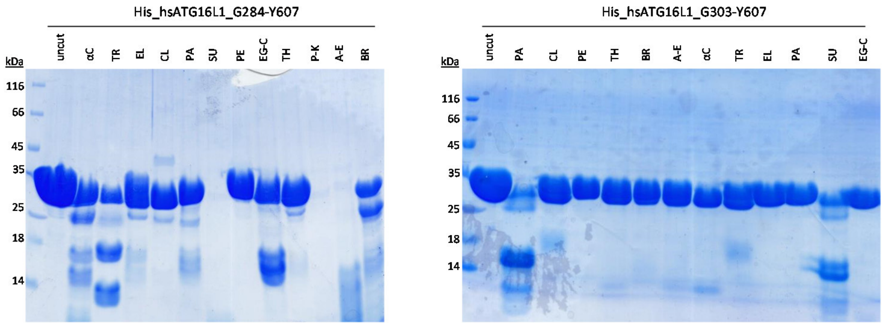


Figure 44: Limited proteolysis of human ATG16L1 G284-Y607 and G303-Y607 constructs. Undigested protein (uncut) was loaded in the first lane as a control. Proteases from Proti Ace Kits 1 and 2 (Hampton Research) were diluted to 0,01 mg/ml and 10 µl of each protease was mixed with 10 µl of each protein (10 mg/ml) and incubated at 37°C for 1 hr. Reaction was stopped by mixing the sample with 2x SDS loading buffer. Samples were analyzed on 15 % SDS PAGE gel. (Legend: α-C: α-chymotrypsin; TR: trypsin; EL: elastase; PA: papain; SU: subtilisin; EG-C: endoproteinase Glu-C; P-K: proteinase K; CL: clostripain (endoproteinase-Arg-C); PE: pepsin; TH: thermolysin; BR: bromelain; A-E: actinase)

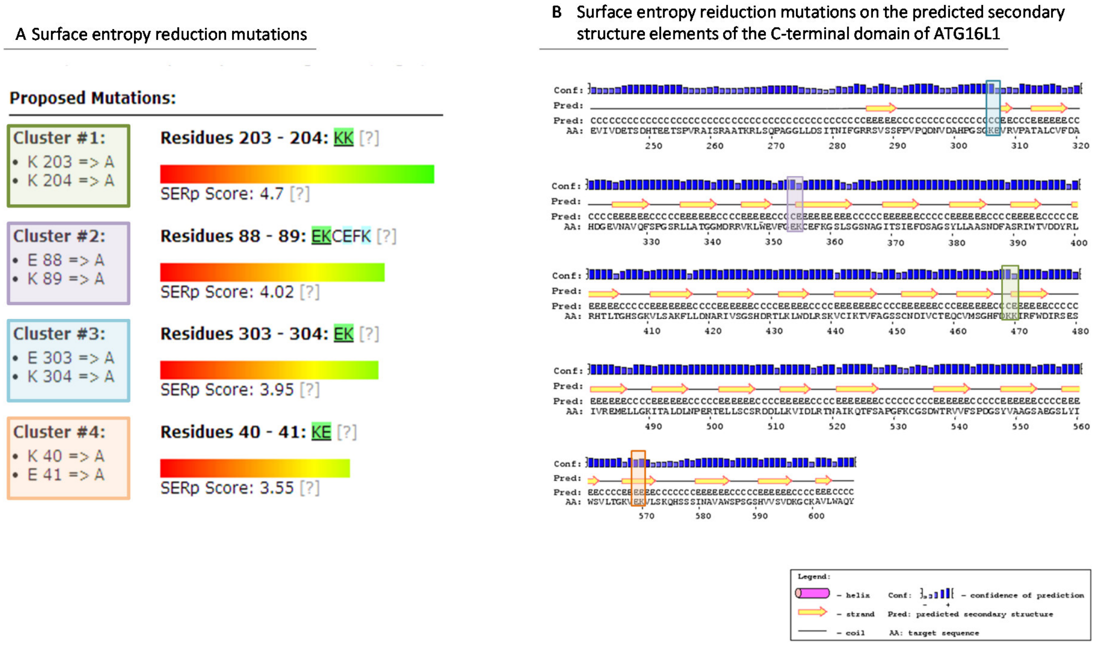
Based on limited proteolysis (Figure 44), ATG16L1_G284-Y607 appears to be more prone to degradation compared to ATG16L1_G303-Y607 construct. Three proteases (trypsin, α-chymotrypsin and bromelain) digest ATG16L1_G284-Y607 into smaller, stable fragments. These three proteases were used for *in-situ* proteolysis (0,01 mg/ml) by adding the proteases to crystallization trails. As cleavage occurs, more stable fragments become available and could separate out of the solution and lead to crystal formation. However, this method did not prove successful for obtaining crystals of the C-terminal domain of human ATG16L1.

Additionally, bromelain produced a shorter, stable fragment of ATG16L1_G284-Y607. This lower band was transferred onto a PVDF nitrocellulose membrane, stained with Coomassie stain and submitted for N-terminal sequencing. Based on the results, the lower fragment starts with position S304, which is very close to the ATG16L1_G303-Y607 construct. Since limited proteolysis of ATG16L1_G303-Y607 was not as prone to proteolysis as ATG16L1_G284-Y607, and the degradation product of ATG16L1_G284-Y607 was one amino acid shorter than ATG16L1_G303-Y607, it would appear the ATG16L1_G303-Y607 represents a compact domain with no major flexible regions.

4.1.8.4 Surface entropy reduction mutagenesis

Another method to modulate already expressed and soluble protein and possibility to increase the chances of crystallizing it, is reducing the entropy of the system. Formation of crystals is dependent on the free energy change, to which enthalpy and entropy contribute. Intramolecular contacts resulting in a crystal are relatively few (around 15 contacts per macromolecule) and the enthalpy values favoring crystal formation are usually small. Thus, protein crystal formation is mainly an entropy driven process. In order for proteins to form intramolecular contacts in a solution, they have to get close enough to each other, resulting in displacement of water (solvent) molecules. Once protein molecules are close enough and crystal contacts are formed, this results in the loss of entropy of protein; compared to the entropy protein has when free in solution. Since formation of crystal contacts releases water molecules normally bound to the protein surface, gain of entropy of the solvent is one of the major driving forces of crystal formation. To promote crystallization through reduction of the entropy of the protein, clusters of flexible, surface exposed residues are mutated to alanine. Surface entropy reduction mutations can facilitate crystal formation in one of two ways: either by reducing the entropy of the protein, a desired effect by introducing SER-mutations, or by simply removing residues with steric hindrance. Changing one residue or a cluster of long, flexible residues might be enough to tilt the balance in favor of crystal formation. SERp server (http://services.mbi.ucla.edu/SER/) predicts secondary structure sequence by PSIPRED and surface conformational entropy based on the submitted amino acid sequence. SERp server predicts patches of surface exposed, charged, flexible residues with high intrinsic entropy which are then mutated to small amino acids with low entropy side chains, such as alanines.

Single predicted surface entropy clusters (Figure 45) were introduced to each of the three C-terminal construct of human ATG16L1 (G266-Y607, G284-Y607 and G303-Y607) for a total of 12 different SER constructs and expressed from HighFive insect cells. Purification procedure for surface entropy mutants was the same as that of the wild type protein. Following gel filtration step, surface entropy mutants of human ATG16L1 were concentrated to 25 mg/ml and either frozen in liquid nitrogen and stored at -80°C, or used for crystallization trails.



A Surface entropy reduction mutations

B Surface entropy reiduction mutations on the predicted secondary structure elements of the C-terminal domain of ATG16L1

Proposed Mutations:

Cluster #1: • K 203 => A • K 204 => A	**Residues 203 - 204:** KK [?] SERp Score: 4.7 [?]
Cluster #2: • E 88 => A • K 89 => A	**Residues 88 - 89:** EKCEFK [?] SERp Score: 4.02 [?]
Cluster #3: • E 303 => A • K 304 => A	**Residues 303 - 304:** EK [?] SERp Score: 3.95 [?]
Cluster #4: • K 40 => A • E 41 => A	**Residues 40 - 41:** KE [?] SERp Score: 3.55 [?]

Figure 45: Patches with high surface entropy predicted for C-terminal domain of human ATG16L1 (R284-Y607) by SERp server. A) SERp output. Amino acid sequence submitted to SERp started at position 266 (which in the prediction is position 1) until 607 (in the prediction position 341). B) Surface entropy reduction mutations predicted by SERp on the secondary structure elements of C-terminal domain of ATG16L1, predicted by PSIPRED v3.3.

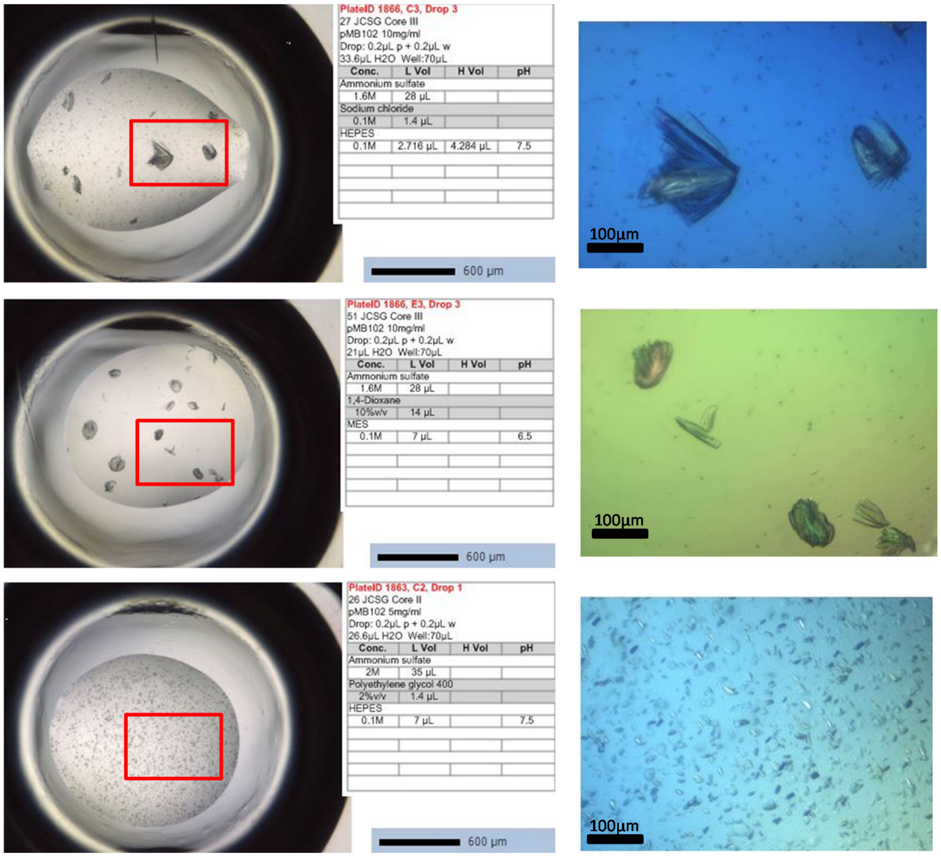
4.1.9 Crystallization of C-terminal domain of human ATG16L1

Crystallization trials of surface entropy reduction variants were set using 5 and 10 mg/ml of protein, with and without tag, at 4°C and 20°C and resulted in many conditions with phase separation. The appearance of phase separation can indicate a metastable transition, a state in which crystals may form. Keeping in mind the concept of small changes of amino acid composition leading to a big change in the crystallization behavior of a macromolecule, naturally optimized constructs (homologues) might provide the change in sequence needed for crystal formation. Human and mouse ATG16L1 (Figure 17) have a few differences at the C-terminal domain. As a start, the most C-terminal human tyrosine 607 was mutated to mouse proline 607 in combination with the surface entropy reduction mutants. Finally, crystals were formed by a construct of human C-terminal domain of ATG16L1 starting at position G303, including surface entropy mutations E354A/K355A along with the Y607P mutation. Initial crystals appeared in three conditions presented in Table 30.

Screen	**Protein Concentration**	**Temp**	**Composition**	**Crystal Morphology**
JSCG Core III	5 and 10 mg/ml	20°C	1.6M $(NH_4)_2SO_4$; 0.1M NaCl; 0.1M HEPES pH 7.5	multiple flower-like cluster of plates
JSCG Core III	5 and 10 mg/ml	20°C	1.6M $(NH_4)_2SO_4$; 10%v/v 1,4-Dioxane; 0.1M MES pH 6.5	multiple flower-like cluster of plates
JSCG Core II	5 mg/ml	20°C	2M $(NH_4)_2SO_4$; 2%v/v PEG400; 0.1M HEPES pH 7.5	Microcrystals

Table 30: Initial crystallization conditions of C-terminal domain of human ATG16L1

Crystals grew in clusters of plates, at both 5 mg/ml and 10 mg/ml at 20°C and first appeared after one or two days and stopped growing after six days. Microcrystals were detected after five to six days under the same conditions as the clusters.



PlateID 1866, C3, Drop 3
27 JCSG Core III
pMB102 10mg/ml
Drop: 0.2µL p + 0.2µL w
33.6µL H2O Well:70µL

Conc.	L Vol	H Vol	pH
Ammonium sulfate			
1.6M	28 µL		
Sodium chloride			
0.1M	1.4 µL		
HEPES			
0.1M	2.716 µL	4.284 µL	7.5

PlateID 1866, E3, Drop 3
51 JCSG Core III
pMB102 10mg/ml
Drop: 0.2µL p + 0.2µL w
21µL H2O Well:70µL

Conc.	L Vol	H Vol	pH
Ammonium sulfate			
1.6M	28 µL		
1,4-Dioxane			
10%v/v	14 µL		
MES			
0.1M	7 µL		6.5

PlateID 1863, C2, Drop 1
26 JCSG Core II
pMB102 5mg/ml
Drop: 0.2µL p + 0.2µL w
26.6µL H2O Well:70µL

Conc.	L Vol	H Vol	pH
Ammonium sulfate			
2M	35 µL		
Polyethylene glycol 400			
2%v/v	1.4 µL		
HEPES			
0.1M	7 µL		7.5

Figure 46: Initial crystals of C-terminal domain of human ATG16L1. Inital drops and conditions are presented on the left hand side. Red areas of interests were imaged under higher resolution and under polarized light and presented on the right hand side.

Initial crystals were reproducible, however always present in a cluster of plates. This might indicate multiple crystal growth from the same origin. However, due to multiplicity of overlapping crystals, collection of a unique diffraction pattern was not possible. Thus, crystals from initial conditions were used to confirm their content. A few of the larger crystals from JSCG Core III C3 and JSCG Core III E3 were harvested, rinsed and dissolved in 2x SDS loading buffer.

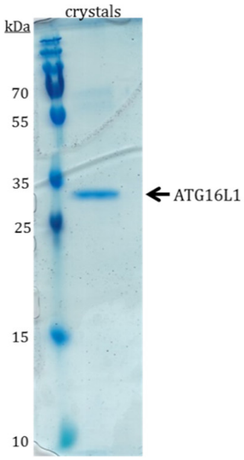


Figure 47: SDS-PAGE analysis of crystals from JSCG Core III C3 and JSCG Core III E3

A single band observed on the gel in Figure 47 corresponds to the size as protein construct used for crystallization trials (ATG16L1_303-607 has a size of 35kDa). The band was sent for tryptic digest and identification of peptides by mass spectrometry. The analysis confirmed that the protein belongs to human ATG16L1 (with protein sequence coverage of 25% and top score of 119).

4.1.9.1 Crystal optimization

The composition of the three conditions that yielded crystal formation is quite similar, with high concentration of ammonium sulfate as main precipitate, in the buffer of pH in the range from 6.5 to 7.5. To optimize the crystal form and obtain single crystals suitable for data collection, random and grid screens were designed taking into account components from the three initial conditions. First round of optimization gave three new conditions at 20°C (Figure 48). There was no crystal growth at 4°C.

Addition of 10 % PEG 3350 to the fine screen caused the crystals to grow slower, appearing only after five days. In addition, it would seem 10 % PEG 3350 reduced the number of nucleation events. In the drops with 10 % PEG 3350, only one to two big crystals were visible, regardless of protein concentration. The condition with 0.1 M NaCl had crystals after two days, with two crystals in the condition with 5 mg/ml of protein, and multiple crystals with 11 mg/ml of protein. Even though many new crystals grew, they were all still multiple, with many crystals growing from the same origin (Figure 48).

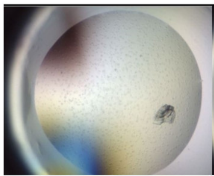
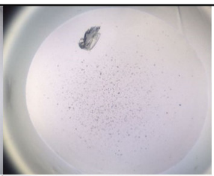
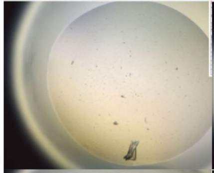
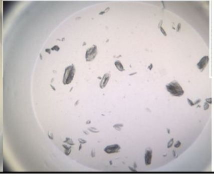
Crystals		**Composition**	**Crystal Morphology**
7 mg/ml	11 mg/ml		
		1.66M $(NH_4)_2SO_4$; 10% PEG 3350; 0.1M MES pH 6.3	phase separation and flower-like cluster of plates
		1.58M $(NH_4)_2SO_4$; 10% PEG 3350; 0.1M MES pH 6.3	phase separation and flower-like cluster of plates
		1.73M $(NH_4)_2SO_4$; 0.117M NaCl; 0.1M MES pH 6.6	phase separation and multiple flower-like cluster of plates

Figure 48: First round of crystal optimization

For the second round of optimization, another random screen was designed with the following ranges:

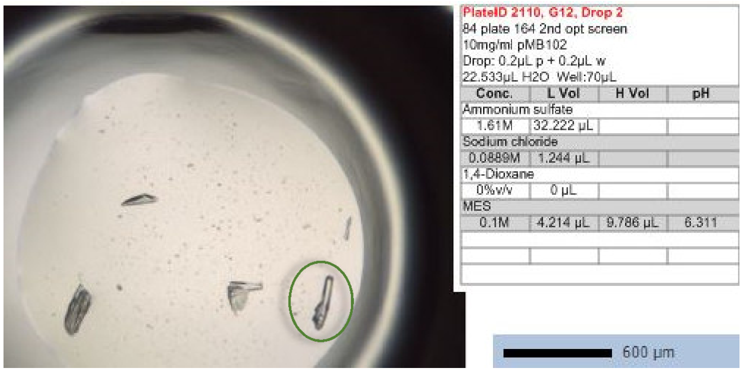
Parameter	Range
pH	6.0-7.5
$(NH_4)_2SO_4$	1.5-2-0 M
PEG 3350	0-10 %
NaCl	0-200 mM
PEG400	0-100 %
1,4-Dioxane	0-5 %
LiCl	0-200 mM

Table 31: List of compounds and their range included in the generation of a random screen

The second optimization screen was set up with 10 mg/ml of protein, in 96-well plates, at 20°C and 30°C, as well as in 15-well plate at 20°C. All the crystals that grew from the second round of optimization were still multiple and clustered. Conditions grown in a 15 well plate did not yield any crystals. Additive screen and silver bullet screens were also tested with the following condition: 1.6 M $(NH_4)_2SO_4$, 10 % PEG 3350, 0.1 M MES pH 6.4. All the crystals grew in clusters.

To overcome clustered crystal growth, some of the crystals from the optimization trials were used to prepare a seed stock and set the first optimization screen with the seed and 3 or 7 mg/ml of protein. The first round of seeding did not yield any single crystals.

The second fine screen set up at 30°C showed some promising crystals. One crystal in particular was a solid three dimensional rectangle on one side, while the other half was a cluster (Figure 49).


PlateID 2110, G12, Drop 2
84 plate 164 2nd opt screen
10mg/ml pMB102
Drop: 0.2µL p + 0.2µL w
22.533µL H2O Well:70µL

Conc.	L Vol	H Vol	pH
Ammonium sulfate			
1.61M	32.222 µL		
Sodium chloride			
0.0889M	1.244 µL		
1,4-Dioxane			
0%v/v	0 µL		
MES			
0.1M	4.214 µL	9.786 µL	6.311

Figure 49: One of the crystals harvested for data collection. Crystal circled in green had a well-structured rectangular half which was used for data collection.

The solid rectangular part of the crystal was frozen in 12 % 2-3 butane-diol as cryoprotectant and measured at BESSY beamline 14.2, Berlin. A data set of 360 degrees was recorded on Pilatus detector. Initial processing was done with XDSAPP. The space group was determined to be centered monoclinic C2, space group number 5. C2 space group has 4 asymmetric units. An asymmetric unit is the smallest part of the unit cell, which can be used to reconstruct the unit cell by applying all the symmetry operations. Thus, one asymmetric unit contains all the information needed to reconstruct the entire crystal.

Wavelength	0.918409 nm
Detector distance	213.97 mm
No. of frames	360
Oscillation range	1 degree/frame
Space group	5 (C2)

Table 32: Data collection

Estimation of the number of macromolecules in the crystallographic asymmetric unit is one of the first steps in structure determination. Matthews coefficient (V_m = volume of asymmetric unit/molecular weight) is directly correlated with the protein crystal solvent content. Based on the unit cell dimensions, space group and molecular weight of the protein, distribution of Matthews coefficient can estimate the number of macromolecules present in the asymmetric unit of the crystal. For the C-terminal domain of ATG16L1, presence of one molecule/AU is most probable with a Matthews coefficient calculated to be 1.90 $Å^3$/Da and solvent content of 35.34 % (Kantardjieff & Rupp 2003; Matthews 1968; Weichenberger & Rup, 2014). The self-rotation function is another tool used for determination of the number of macromolecules in the asymmetric unit and how they are related. Figure 50 depicts the self-rotation function calculated using the data set of C-terminal domain of ATG16L1 and the two strongest peaks at 90°/-90° indicate the twofold crystallographic rotation axis.

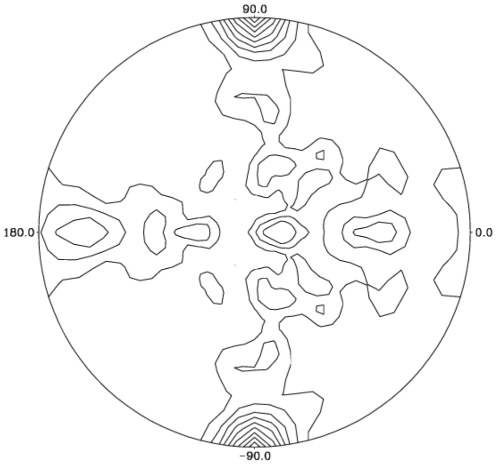


Figure 50: Self rotation function. The κ = 180° section of the self-rotation function calculated for the ATG16L1_303-607 data set using POLARRFN (Collaborative Computational Project No. 4, 1994) *with an integration radius of 20 Å and data in the resolution range 20–4 Å.*

4.1.9.2 Structure determination

Structure of C-terminal domain of ATG16L1 was solved using a molecular replacement approach. This method relies on the existence of already known protein structures which are similar to the target protein. The most well-known algorithm that allows comparison of the protein sequence against a non-redundant database of all known sequences is BL AST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Peptide sequence for C-terminal domain of human ATG16L1 (303-607) was submitted for BLAST alignment against the PDB data base and resulted in the following list of hits:

Database: PDB protein database

77,857 sequences; 19,029,097 total letters

Query Length=305

Sequences producing significant alignments	**Score (Bits)**	**E Value**
pdb|3N0E|A Chain A, Crystal Structure Of Wdr5 Mutant (W330y)	108	5e-27
pdb|3N0D|A Chain A, Crystal Structure Of Wdr5 Mutant (W330f)	108	6e-27
pdb|3MXX|A Chain A, Crystal Structure Of Wdr5 Mutant (S62a)	108	8e-27
pdb|2CNX|A Chain A, Wdr5 And Histone H3 Lysine 4	107	2e-26
pdb|2YMU|A Chain A, Structure Of A Highly Repetitive Propeller	109	2e-26
pdb|2CO0|A Chain A, Wdr5 And Unmodified Histone H3 Complex	107	3e-26
pdb|2G99|A Chain A, Structural Basis For The Specific Recognition	105	6e-26
pdb|2H68|A Chain A, Histone H3 Recognition And Presentation By Wdr5	105	6e-26
pdb|3SMR|A Chain A, Crystal Structure Of Human Wd Repeat Domain	105	6e-26
pdb|2H13|A Chain A, Crystal Structure Of Wdr5 histone H3 complex	105	6e-26

Table 33: BLAST results of C-terminal domain of ATG16L1 against the PDB data base. Only first top hits shown.

According to BLAST search, ATG16L1 G303-Y607 has highest similarity with WDR5 (PDB code 2CNX) with sequence identity of 27 % and sequence similarity of 51 %. The structure of 2CNX was submitted to CHAINSAW (Stein 2008) from CCP4, a program which aligns the target and model sequence and modifies the side chains of the model in one of three ways: either by reducing all the side chains to the Cβ, or by reducing all the side chains to Cγ, or it preserves all the atoms the target and the model have in common. Since WDR5 and ATG16L1 are not homologs, CHAINSAW reduced the side chains of the WDR5 to the Cβ, thus generating a poly-alanine backbone from the 2CNX PDB file. The poly-alanine model was then submitted to Phaser-MR in Phenix (**P**ython-based **H**ierarchical **En**vironment for **I**ntegrated **X**tallography (Adams *et al.* 2010)). Placement of the search PDB poly-alanine model resulted in Top LLG (log likelihood gain) of 122.297, which indicated the molecular replacement was successful, but the side chain electron density might be difficult to interpret (Figure 51). SHELXE (Sheldrick 2007; Sheldrick 2002) is another program used for improvement of molecular replacement solutions. SHELXE was originally designed for experimental phasing using MAD/SAD/SIRAS data, by providing iterative phase improvement by density modification. This same approach can be used to improve phases of the poly-alanine model using phases calculated from the molecular replacement model rather than experimental phases. SHELXE uses calculated phases from the model with experimental, measured intensities as Fourier coefficients for calculation of the electron density. Poly-alanine chain from the model is then built in the electron density, which undergoes rounds of refinement to reduce model bias. SHELXE resulted in a map with more clearly defined electron density for side chains of the target protein. The difference between initial electron density maps calculated by Phaser-MR and SHELXE is represented in Figure 51. Bulky and/or characteristic residues of the target protein are chosen to demonstrate the difference between Phaser-MR and SHELXE initial electron density solutions.

Output file from SHELXE is further refined in Phenix. Model building was done in COOT (Emsley *et al.* 2004; Emsley *et al.* 2010), an application for model building and validation of macromolecular structures. COOT displays an electron density map and atomic models and allows its manipulation through the use of various optimization algorithms. Structure of ATG16L1_303-607 had a defined density and could be built from amino acid 307 to 607. The N-terminal part, namely residues corresponding to the His tag, TEV cleavage site, and to a short stretch from position 303 to 306 were not visible in the electron density map. Refinement statistics are summarized in Table 34.

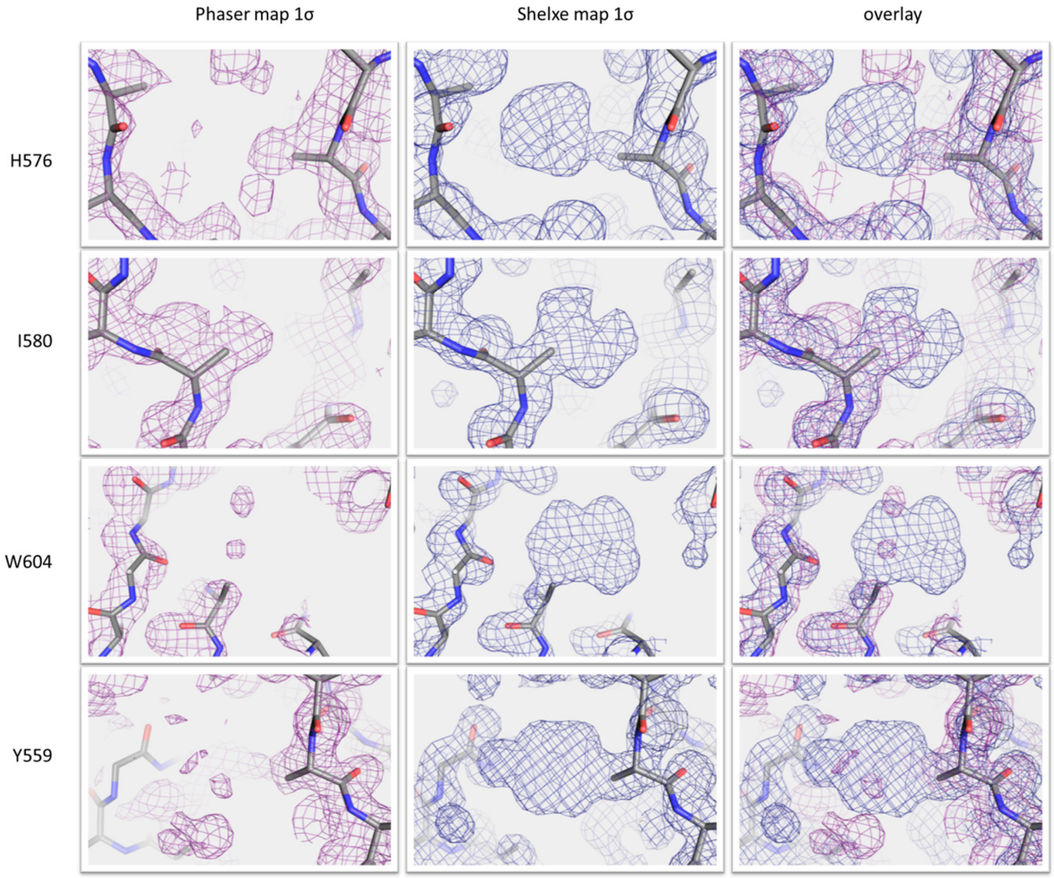


Figure 51: Initial electron density map with poly-alanine model. Light pink is electron density map calculated by Phaser-MR: Blue is electron density map calculated by SHELXE. Third set of panels represents the overlay of two maps. Residues of the target protein corresponding to each panel set are listed on the left. All maps are contoured at 1σ level. Atoms of the poly-alanine chain are colored according to the following scheme: carbon in gray, nitrogen in blue and oxygen in red.

Space group	C 2
Unit cell dimensions *a*, *b*, *c* (Å)	158.1 38.3 45.8
Unit cell angles	90.0 105.7 90.0
Resolution (Å)	50.0 - 1.7 (1.8 - 1.7)
Measured reflections	198313 (30838)
Unique reflections	29397 (4553)
I/σ	20.8 (4.8)
Completeness (%)	99.8 (99.2)
Redundancy	6.7
Wilson *B* factor (Å^2)	17.46
R factor observed (%)	6.1 (39.2)
R factor expected (%)	6.1 (38.7)
R-meas (%)	6.6 (42.5)
Solvent content (%)	35.34
R_{work} (%)	16.82
R_{free} (%)	19.74
Protein atoms	2323
Solvent atoms	144
r.m.s.d. from ideal geometry:	
-Bond lengths (Å)	0.006
-Bond angles (°)	1.13
Average *B* factor (Å^2):	26.75
Ramachandran plot (%):	
-Favored region	95.5%
-Allowed region	3.8%
-Outliers	0.7%

Table 34: Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

Since phases are either experimentally determined by SIR/MIR or SAD/MAD methods, or inferred by molecular replacement, the quality of macromolecular structure has to be judged by several parameters to ensure the integrity of the data and minimize the amount of errors introduced during model building and refinement process. The program POLYGON (Urzhumtseva *et al.* 2009) compares different indicators of model quality with a number of PDB structures with similar resolution. Quality criteria, such as R factors and average *B* factor, are also presented in the refinement statistics in Table 34. However, POLYGON compares these values together with three additional quality indicators (r.m.s.d. for bonds and angles, and a clash score) with values calculated for 824 different structures at the same resolution. Analysis of structural parameters such as bond length and bond angles provide information about the quality of the structure. For a high quality structure, r.m.s.d. values for bond lengths should be below 0.02 Å and below 4 degrees for bond angles. Clashes are defined as non-covalent unfavorable contacts. Value for the clash score should be as low as possible, though a clash score of 0 is unusual. Values for the target structure are plotted as a point on a ruler, and connected to with a black line to form a polygon. POLYGON plot for the structural model of C-terminal domain of human ATG16L1 calculated during this work is presented in Figure 52.

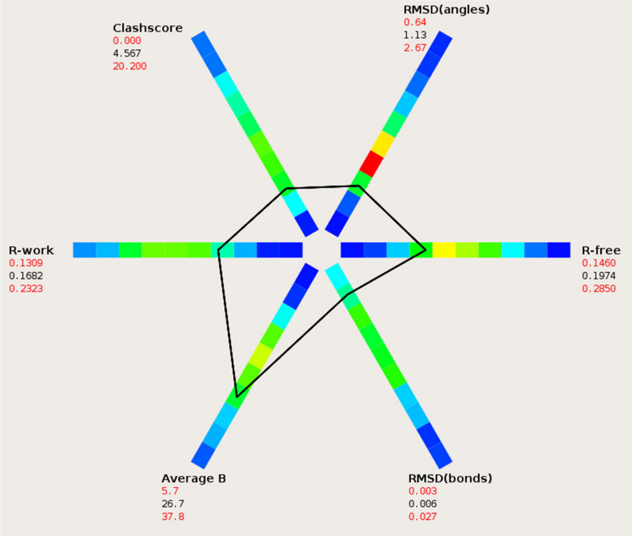


Figure 52: Polygon plot by Phenix. Polygon shows histograms of the distribution of selected statistics across 824 PBD entries of similar resolution, with the range specified by numbers in red. Statistics for the structure of C-terminal domain of human ATG16L1 are printed in black. Frequency of observed values in the reference set of structures is color coded: red represents low frequency, blue represents high frequency)

Three torsion angles define the configuration of a polypeptide backbone. The peptide bond, with its partial double bond character, can have torsion angle omega (ω) values of 0° or 180°. The other two torsion angles, phi (ϕ) and psi (ψ) are constrained by steric clashes within each isolated amino acid and handedness of the amino acids. Favored combinations of conformational angles (phi and psi) for each residue along the peptide chain are greatly restricted and described by a Ramachandran plot. Torsion angles phi and psi define either side of the Cα of each residue: phi is defined by N-Cα bond, while psi is defined by Cα-C bond. Ramachandran plot reveals plausibility of the geometry of the polypeptide backbone and is used as tool for validation of structures.

Typically, residues involved in β-sheet formation are in the preferred upper left region of the plot. For the crystal structure of C-terminal domain of human ATG16L1, 99% of all amino acids fall into favored and additionally allowed regions of the Ramachandran plot (Figure 53).

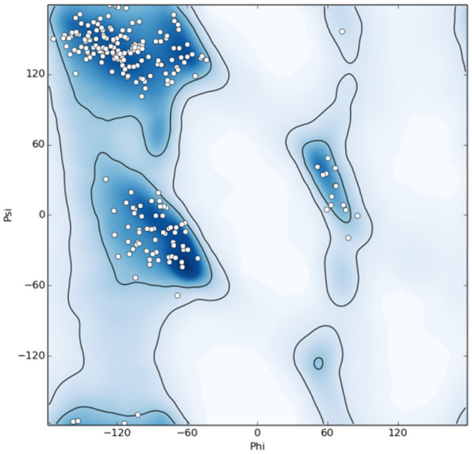


Figure 53: Ramachandran plot for all non-proline and glycine residues of C-terminal domain of human ATG16L1. Dark blue regions represent torsion angle combinations that give no steric hindrance, while light blue areas result in modest repulsion.

4.1.9.3 Structure of C-terminal domain of human ATG16L1

X-ray crystal structure of C-terminal domain of ATG16L1 (303-607) was solved by molecular replacement with a human WDR5 structure as a model. WDR5 is a WD repeat containing protein involved in histone modification (Ruthenburg *et al.* 2006). Crystal structure of the C-terminal domain of ATG16L1 (303-607) confirms this domain folds into a WD-40 propeller, as predicted by Phyre2 (Figure 15). WD-40 domain of ATG16L1 has a characteristic donut shape with seven β-sheets, each composed of four anti-parallel β-strands. From the side view, a narrow top side and a wider lower side is observed (Figure 54).

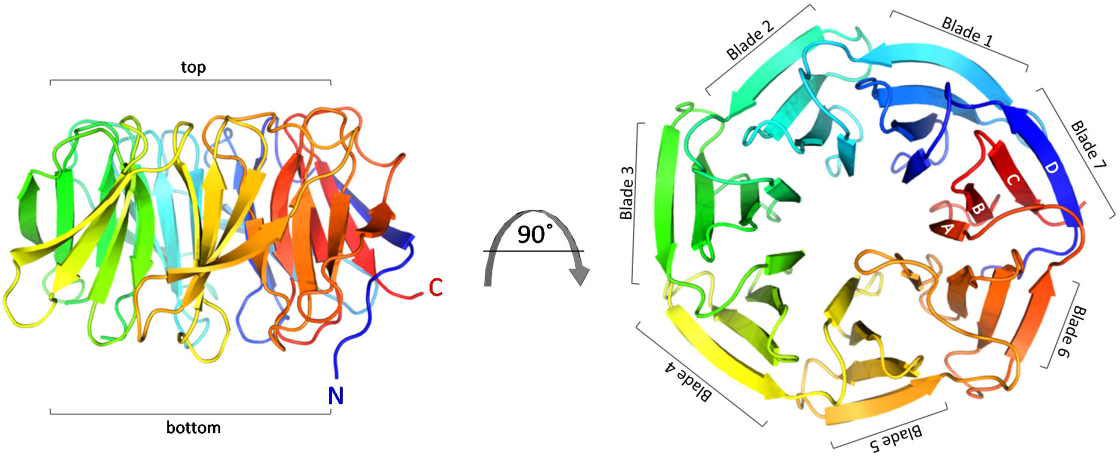


Figure 54: Crystal structure of the C-terminal domain of human ATG16L1 in cartoon representation. The peptide chain is colored according to rainbow scheme from the N-terminus (blue) to C-terminus (red).

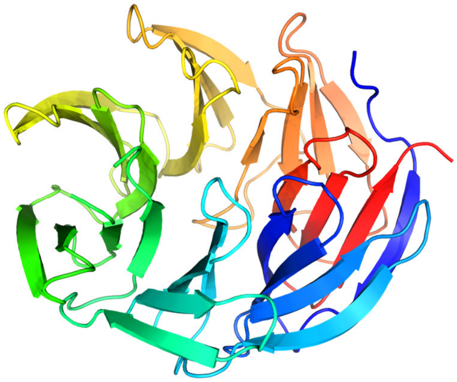


Figure 55: Stereo view of the WD-40 domain of human ATG16L1. The peptide chain is colored according to rainbow scheme from the N-terminus (blue) to C-terminus (red)

The model shown in Figure 54 and Figure 55 is based on the data set with a maximum resolution of 1.7 Å. Structures in the 1.2-2.0 Å resolution range are considered high resolution structures since the level of detail is good enough to clearly distinguish between different amino acid side chains. This can be observed in the selection of electron density frames from the model of the WD40 of ATG16L1 displayed in Figure 56.

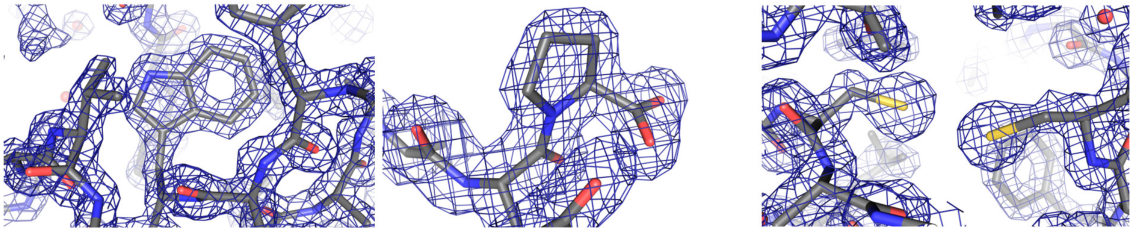


Figure 56: 2mFo-DFc maps. First panel shows a tryptophan residue at 2σ level. Second panel shows a proline residue with carboxyl terminus at 1σ level. Third panel shows two cysteine residues at 2σ level.

4.1.9.4 Crystal packing

C-terminal domain of human ATG16L1 crystalized in centered monoclinic C2 with one monomer in the asymmetric unit.

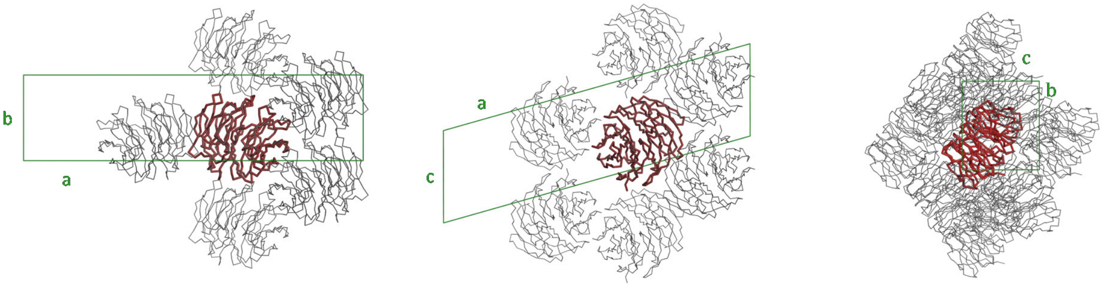


Figure 57: Crystal packing of crystals of C-terminal domain of ATG16L1. Symmetry-related molecules are shown in gray.

5 Discussion

5.1 WD-40 proteins

Most proteins are composed of one or more domains, building blocks of protein structures. A domain is considered an independent structural and functional unit of a protein. Some domains are well characterized and specific, and their presence in an unknown protein strongly suggests the function of the protein. In other cases, the same domain can have multiple, unrelated functions depending on the protein it belongs to. One such domain with functional diversity is the WD-40 domain. Proteins containing WD-40 domain are widely present in eukaryotes and are involved in RNA processing, signal transduction, gene regulation and development, vesicular transport, cell cycle control, apoptosis and chromatin dynamics (Neer *et al.* 1994; Xu and Min 2011). The main common function of all WD-40 proteins described so far is to facilitate reversible interactions with other proteins, peptides or nucleic acids. WD-40 domains have also been reported to catalyze enzymatic reactions (Suzuki *et al.* 2005) adding to the ever growing repertoire of functions described for this domain. Based on secondary structure predictions of the primary sequence, C-terminal half of human autophagy core protein ATG16L1 is a WD-40 propeller domain (Figure 14).

WD-40 domain comprises of WD-40 repeats, highly conserved in eukaryotes, with a characteristic sequence and structural features. A WD-40 repeat was first described as a reiterated pattern of a 43 amino acid long β-sheet with highly conserved arginine (R) at position 6, glycine-histidine (GH) dipeptide at position 13 and 14, aspartic acid (D) at position 36 and a terminal tryptophan-aspartic acid dipeptide (WD) at positions 42 and 43 (Fong *et al.* 1986). This definition was further refined to define the WD-40 repeat as a peptide of 44-60 amino acids, with conserved GH dipeptide at 11-24 position and a C-terminal WD (Neer *et al.* 1994; Smith *et al.* 1999). The name of the motif is based on the terminal WD dipeptide, and an approximated initial length of around 40 amino acids.

Secondary structure of a WD-40 repeat is an anti-parallel β-sheet composed of four β-strands (Figure 58(iii)). A number (typically seven) of anti-parallel β-sheets fold together into a doughnut-like β-propeller scaffold structure (Figure 58(i and ii)). A single anti-parallel β-sheet is commonly referred to as a blade, because it resembles a blade in a propeller. Using X-ray crystallography methods, a high resolution structural model of the C-terminal domain of human ATG16L1 confirmed the WD-40 propeller of human ATG16L1 folds into a typical WD-40 domain, with seven β-sheets (blades), each composed of four anti-parallel β-strands (Figure 54).

A WD-40 propeller has three large surfaces: a narrow top, a wider, bottom surface and the entire circumference of the doughnut (Figure 58(i)). Top surface is defined as the side with the loop connecting strand D from one blade with strand A from the next (Smith *et al.* 1999; C. Xu & Min, 2011). One of the arguments why WD-40 domains are so conserved and versatile at the same time comes exactly from the great interaction platform this domain provides.

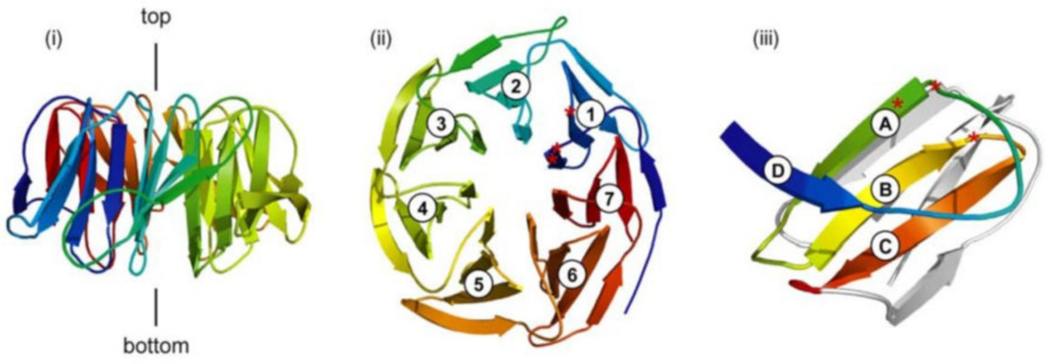


Figure 58: Overall WD-40 domain architecture. (i) Side view of the WD-40 propeller. (ii) Top view of the WD-40 propeller with 7 anti-parallel β-sheets numbered. Each β-sheet is composed of 4 β-strands. (iii) Structural arrangement of one WD-40 repeat, formed by three (A-C) β strands of one, and the forth β-strand (D) from another β-sheet. (Stirnimann *et al.*, 2010)

Such arrangement, where each blade is composed of three strands of one WD-40 repeat and the last for the next, provides stability and rigidity to the overall domain. This characteristic arrangement is also present in the WD-40 domain of ATG16L1 (Figure 54).

Each blade of the typical seven-bladed WD-40 domain has two main variable regions, which are not essential for the fold of the domain, but more likely provide the specificity and the function. The largest flexible region, represented in green in Figure 59, is between strand C of one blade and strand A of the next (including strand D). This region can vary between 11 and 150 residues in length. The loop between strands C and D protrudes from the bottom of the propeller, while the loop between strands D and A of the next blade is exposed to the top surface. Strand D is exposed to the outside surface of the propeller. Second variable region (Figure 59, in red) is shorter, between strands A and B of the same blade, and forms a loop exposed to the bottom of the propeller (Smith *et al.* 1999).

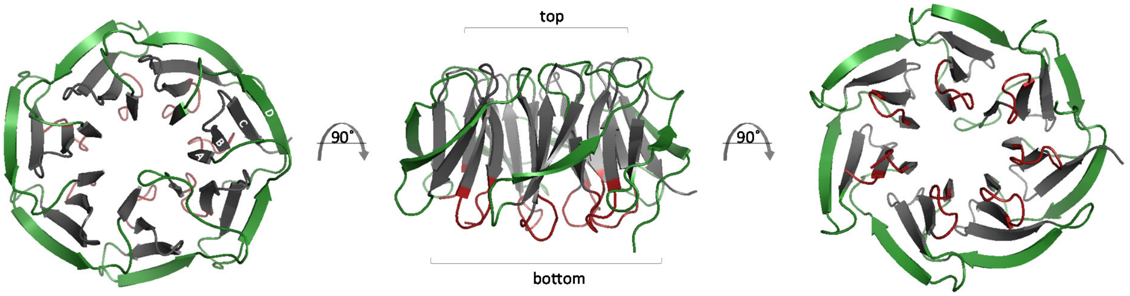


Figure 59: Two flexible regions of WD-40 domain. Typical, seven bladed WD-40 domain is represented in gray. The largest flexible region between strand C of one blade and strand A of the next is colored in green. Second variable region between strands A and B of the same blade is colored in red.

Due partially to these variable regions, interaction surface of the WD-40 proteins with other proteins and peptides are more often at the top, narrow end (Russell *et al.* 1998). However, the bottom surface of the propeller with the variable region between strands C and D as well as the circumference composed mainly of un-conserved D strands have also been shown to provide a platform for protein binding (Figure 60). The variability in the regions of binding areas might be another reason why these domains are so common. Creating a new binding site by introducing changes to one or two WD-40 repeats, while preserving the common fold, might be evolutionarily more efficient than generating an insertion or deletion which might lead to drastic changes in the structure of other domain families, lacking the regular repeats to maintain the overall domain fold. (Stirnimann *et al.* 2010).

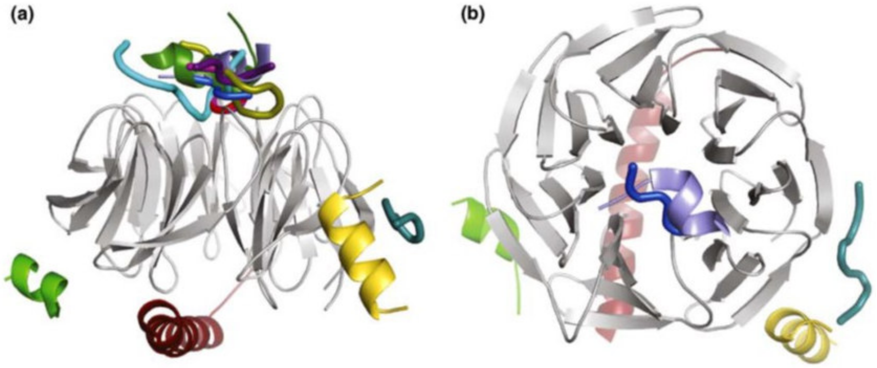


Figure 60: Interaction surfaces of WD-40 domain. (a) Side view of the WD-40 domain showing 13 known interaction partners. Most of the interactions are concentrated on the top surface, close to the center. (b) Top view of the WD-40 domain showing 6 known interaction partners occupying all three possible binding surfaces: top, bottom and the circumference. (Stirnimann *et al.*, 2010)

5.2 WD-40 domain of human ATG16L1

5.2.1 Conservation of WD-40 domain of ATG16L1 in higher eukaryotes

A key member of core autophagy machinery involved in nucleation and elongation of the phagophore membrane is ATG16L1. In yeast, this protein is 150 amino acids long and composed of an ATG5 binding domain and a coiled coil domain. This basic architecture present in the yeast ATG16 protein is conserved through eukaryotic systems and essential for canonical autophagy. In addition to these two fundamental domains, ATG16 in higher eukaryotes additionally has a WD-40 domain, absent from the yeast homologue (Figure 61).

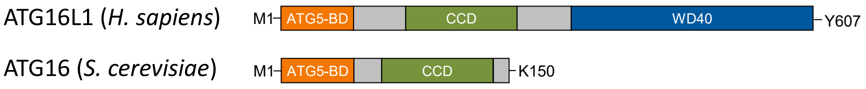


Figure 61: Graphical domain representation of human ATG16L1 and yeast ATG16

Absence of the WD-40 domain from the yeast ATG16 indicates this domain is not involved in canonical, bulk autophagy induced by starvation or rapamycin treatment. Comparison of human ATG16L1 protein sequence across species shows WD-40 domain is well conserved in higher eukaryotes. This cross-species conservation of the WD-40 domain indicates certain structural and functional importance. WD-40 domain of human ATG16L1 (303-607) was used in BLAST-p to search for non-redundant orthologues. A broad range of orthologues was detected including invertebrates, amphibians, fish, birds and mammals. An alignment of WD-40 domain of ATG16L1 from five representative species was performed using Clustal Omega and the result is presented in Figure 62.

```
X.      GNREVRVPATAVYSFDAHDGEVNAVRFSPGSRLLATGGMDRRVKLWDVIGNKCEAKGSLTGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDNARI    120
H.      SGKEVRVPATALCVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEVFGEKCEFKGSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDNARI    120
M.      SGKDVRVPTTASYVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEAFGDKCEFKGSLSGSNAGITSIEFDSAGAYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDNARI    120
A.      ASKEVRVPTTAICVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEVLGDRCEPKGSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDNRLRHTLTGHSGKVLSAKFLLDNARI    120
C.      SSKEVRVPTTAVYVFDAHDGEVNAVQFSPASRLLATGGMDRRVKLWEVFGDRCELKSSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDNARI    120
        . ::****:**   ***********:***.****************:.:*::** *.**:***************:****************** *************************

X.      VSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIICTEQCVMSGHFDKKIRFWDIRTECIVRELELQGRITALDLNPERTQLLSCSRDDLIKITDLRANAVQQTFSAPGFKCGSDWTRVIFS    240
H.      VSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSESIVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAIKQTFSAPGFKCGSDWTRVVFS    240
M.      VSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRSESVVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAVKQTFSAPGFKCGSDWTRVVFS    240
A.      VSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRTESIVKELELLGRITALDLNSERTELLTCSRDDLLKIIDLRVSAVKQTFSAQGFKCGSDWTRVVFS    240
C.      VSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTEQCVMSGHFDKKIRFWDIRTESIVQELELFGRITALDLNPERTELLTCSRDDLLKIIDLRVNAVKQT----------------FS    224
        ********************************:*********************:*.:*:*:** *:******* ***:**:******:*: ***..*::**                **

X.      PDGNYVSAGSAEGTLYFWNVLTGKMERMFSKQHSSSINAVAWSPSGTHVVSVDRGSRGVLWSDF    304
H.      PDGSYVAAGSAEGSLYIWSVLTGKVEKVLSKQHSSSINAVAWSPSGSHVVSVDKGCKAVLWAQY    304
M.      PDGSYVAAGSAEGSLYVWSVLTGKVEKVLSKQHSSSINAVAWAPSGLHVVSVDKGSRAVLWAQP    304
A.      PDGSYVAAGSADGALYIWSVLTGKLERTLAKHHSSSINAVAWSPAGAHVVSVDKGNKAVLWSEF    304
C.      PDGSYVAGGSADGVLYIWNVLTGKIERTLAKHHSSSINAVAWSPAGAHVVSVDKGNKAVLWSEF    288
        ***.**:.***:* **.*.*****:*: ::*:**********:*:* ******:* :.***::
```

*Figure 62: Clustal Omega multiple alignment of the WD-40 domain of ATG16L1 in higher eukaryotic organisms. X. = Xenopus laevis (African clawed toad), H. = Homo sapiens (human), M. = Mus musculus (house mouse), A. = Acanthisitta chloris (small bird), C = Chelonia mydas (green sea turtle). An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties.*

Human WD-40 domain of ATG16L1 has 95% sequence identity with the mouse orthologue, 89% with *A. chloris* (a small bird), 86% with *X. laevis* (African clawed toad) and 84% with *C. mydas* (green sea turtle). Amino acids marked with an asterisk (*) in Figure 62 are conserved across higher eukaryotic species. On the structural model of WD-40 domain of ATG16L1, these conserved residues are marked in orange (Figure 63) and are distributed throughout the WD-40 domain of ATG16L1.

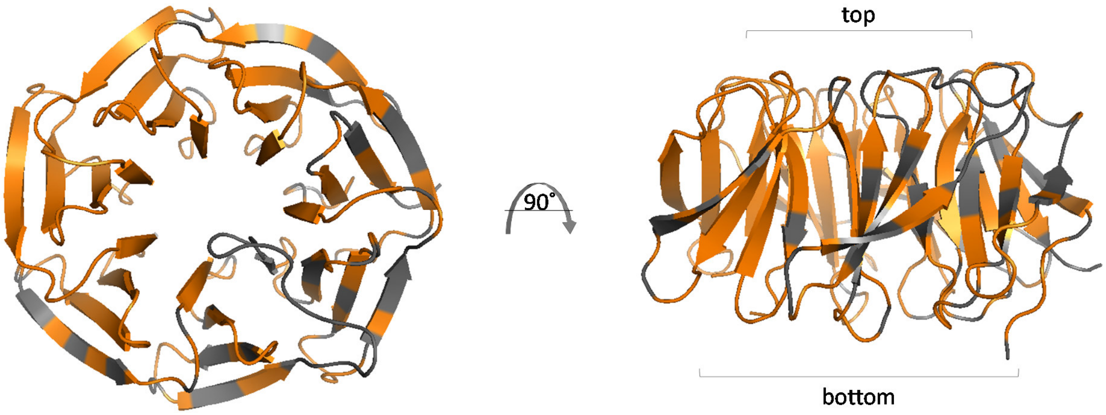
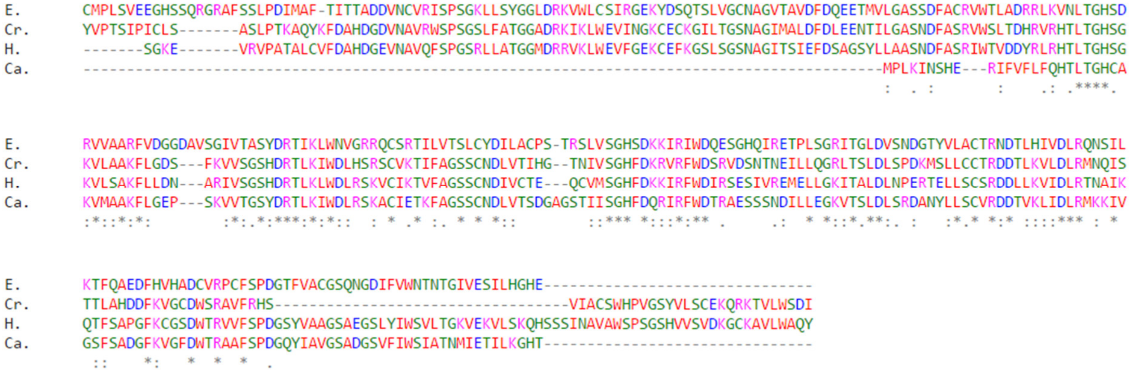


*Figure 63: Conserved WD-40 domain residues. Residues marked with * (asterisk) in Figure 62 are marked in orange on the WD-40 domain of ATG16L1 (represented in cartoon, in gray).*

In the case of lower eukaryotes, both sequence similarity and conservation in this region of the protein are low (Figure 64). Human WD-40 domain (303-607) has 56 % sequence identity with *Camponotus floridanus*, 51 % with *Crassostrea gigas*, and 45 % with *Echinococcus granulosus*.



```
E.      CMPLSVEEGHSSQRGRAFSSLPDIMAF-TITTADDVNCVRISPSGKLLSYGGLDRKVWLCSIRGEKYDSQTSLVGCNAGVTAVDFDQEETMVLGASSDFACRVWTLADRRLKVNLTGHSD
Cr.     YVPTSIPICLS-------ASLPTKAQYKFDAHDGDVNAVRWSPSGSLFATGGADRKIKLWEVINGKCECKGILTGSNAGIMALDFDLEENTILGASNDFASRVWSLTDHRVRHTLTGHSG
H.      -------SGKE-------VRVPATALCVFDAHDGEVNAVQFSPGSRLLATGGMDRRVKLWEVFGEKCEFKGSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSG
Ca.     --------------------------------------------------------------------------------------------MPLKINSHE---RIFVFLFQHTLTGHCA
                                                                                                    :  . :       :    .: .****.

E.      RVVAARFVDGGDAVSGIVTASYDRTIKLWNVGRRQCSRTILVTSLCYDILACPS-TRSLVSGHSDKKIRIWDQESGHQIRETPLSGRITGLDVSNDGTYVLACTRNDTLHIVDLRQNSIL
Cr.     KVLAAKFLGDS---FKVVSGSHDRTLKIWDLHSRSCVKTIFAGSSCNDLVTIHG--TNIVSGHFDKRVRFWDSRVDSNTNEILLQGRLTSLDLSPDKMSLLCCTRDDTLKVLDLRMNQIS
H.      KVLSAKFLLDN---ARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVCTE---QCVMSGHFDKKIRFWDIRSESIVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAIK
Ca.     KVMAAKFLGEP---SKVVTGSYDRTLKIWDLRSKACIETKFAGSSCNDLVTSDGAGSTIISGHFDQRIRFWDTRAESSSNDILLEGKVTSLDLSRDANYLLSCVRDDTVKLIDLRMKKIV
        :*::*:*:        :*:.*:***:*:*::  : * .* :. * * *::        ::*** *:::*:** .     .:  * *::*.**:. :   :*.* *:* ::::*** : *

E.      KTFQAEDFHVHADCVRPCFSPDGTFVACGSQNGDIFVWNTNTGIVESILHGHE-------------------------------
Cr.     TTLAHDDFKVGCDWSRAVFRHS----------------------------------VIACSWHPVGSYVLSCEKQRKTVLWSDI
H.      QTFSAPGFKCGSDWTRVVFSPDGSYVAAGSAEGSLYIWSVLTGKVEKVLSKQHSSSINAVAWSPSGSHVVSVDKGCKAVLWAQY
Ca.     GSFSADGFKVGFDWTRAAFSPDGQYIAVGSADGSVFIWSIATNMIETILKGHT-------------------------------
         ::    *:   *  *  *  .
```

*Figure 64: Clustal Omega multiple alignment of the WD-40 domain of ATG16L1 with lower eukaryotic organisms. E. = Echinococcus granulosus (type of tape worm), Cr. = Crassostrea gigas (oyster), H. = Homo sapiens (human), Ca. = Camponotus floridanus (carpenter ant). An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties*

Naturally, the sequence identity between human and low eukaryotic organisms is very low. However, the ATG16L1 of tape worm and oyster have what seems to be a more complete WD-40 domain compared to the carpenter ant. The fact that the WD-40 domain is missing from yeast ATG16 but is partly present in other lower eukaryotes and conserved in higher eukaryotes indicates that the WD-40 domain of ATG16L1 evolved with different species, and is possibly still evolving, adjusting itself to the unique requirements and needs of organisms it belongs to and their environments.

5.2.2 Comparison between WD-40 domain of ATG16L1 and other WD-40 domains

WD-40 domain proteins are one of the largest domain families present in eukaryotes with the widest range of interaction partners (Stirnimann *et al.*, 2010). WD-40 domain can interact with other proteins, peptides and nucleic acids using one of three large surfaces (top, bottom and circumference) (Stirnimann *et al.*, 2010). Most structurally characterized WD-40 domains have seven blades and secondary structure prediction suggests the WD-40 domain of human ATG16L1 is composed of 30 β-strands (Figure 14). Furthermore, Hampe *et al.* proposed this domain to have eight C-terminal repeats with 32 β-strands based on extensive use of alignment and prediction algorithms together with already structurally characterized WD-40 domains (Hampe *et al.* 2006). However, the high resolution X-ray structure of WD-40 propeller of ATG16L1 revealed during this work clearly shows the complete propeller is composed of seven blades, each comprising of four anti-parallel β-strands (resulting in a total of 28 β-strands, Figure 54). As already mentioned, due to low sequence conservation and high functional diversity of WD-40 domains, results obtained by prediction algorithms must be treated with great caution. They are very powerful and their importance in crystallography and protein study should not be underestimated, however their shortcomings must also be considered. In the case of the WD-40 domain of ATG16L1, and possibly in many other cases, prediction and modeling studies have been enough to predict the overall fold of the domain, but only high resolution structural determination could show the amount of detail needed for more in depth understanding of this promiscuous domain.

To determine the structure of human ATG16L1 WD-40 domain with molecular replacement method, WD-40 domain of ATG16L1 (303-607) was used in BLAST-p to search for structurally similar models in the protein data base (PDB). From the results, top five structures with highest alignment score are listed in Table 35.

PDB code	Protein	Organism	Cellular process	PyMOL r.m.s.d. (Å)	Max Score	Seq ident (%)
2CNX_A	Wdr5	*Homo sapiens*	transcription	1.1	107	27
4WJV_A	Rsa4	*Saccharomyces cerevisiae*	ribosome biogenesis	7.5	89.4	26
1VYH_C	Paf-Ah	*Mus musculus*	signal transduction	1.5	85.9	23
2OVP_B	Skp-1	*Homo sapiens*	cell cycle	3.2	79.3	24
4LG9_A	Tbl1xr1	*Homo sapiens*	unknown	2.8	77.8	24

Table 35: List of top five structurally characterized WD-40 domains identified used BLAST-p and PDB. The root-mean-square deviation (r.m.s.d.) is the measure of the average distance between the Cα atoms of superimposed proteins. Max score described the highest alignment score between the query sequence and the database sequence segment.

The result with the highest score, Wdr5 (PDB: 2CNX_A, (Ruthenburg *et al.*, 2006)) was successfully used as a molecular replacement model for structure determination of WD-40 domain of ATG16L1. Superposition of WD-40 domain of ATG16L1 with the top hit from Table 35, Wdr5, (Figure 65) resulted in lower r.m.s.d. value compared to other structures, which confirms the assumption that based on current structural knowledge of WD-40 domains, Wdr5 is the best fitting model for WD-40 domain of ATG16L1.

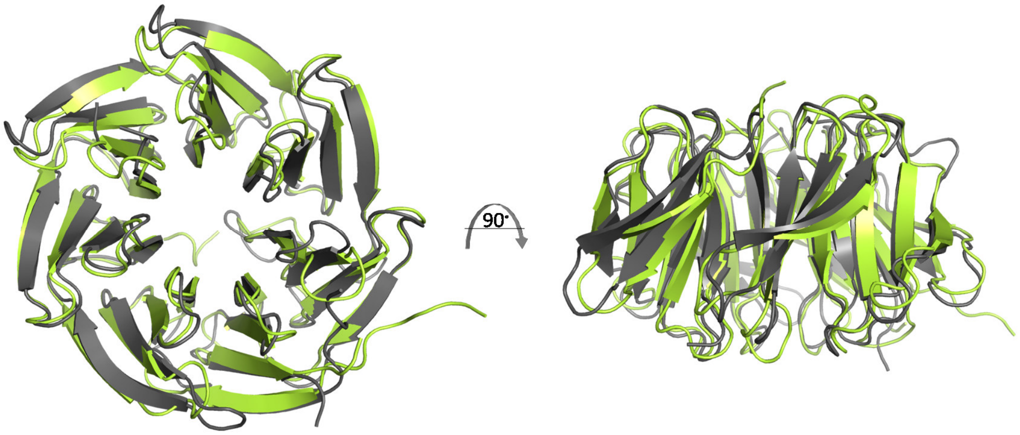


Figure 65: Superposition of the WD-40 domain of ATG16L1 (in gray) with Wdr5 (PDB: 2CNX, (Ruthenburg *et al.*, 2006)*, in lemon).*

5.2.3 Discrepancies between predicted and actual secondary structure elements of WD-40 domain of ATG16L1

Since not all WD-40 propeller domains have seven blades, and due to the low sequence conservation of the domain, prediction algorithms often omit or add one or two β-strands at the beginning or at end of the domain. However, without a structural model, these predictions are the best information available. Since structural studies rely on the predictions of secondary structures for construct design, it is also essential that the shortcomings of these predictions be understood and considered during determination of construct boundaries. Once the structural model is available, true arrangement and architecture of the domain is evident, and can then be compared to the predicted elements. For the WD-40 propeller of human ATG16L1 this comparison is presented in Figure 66.

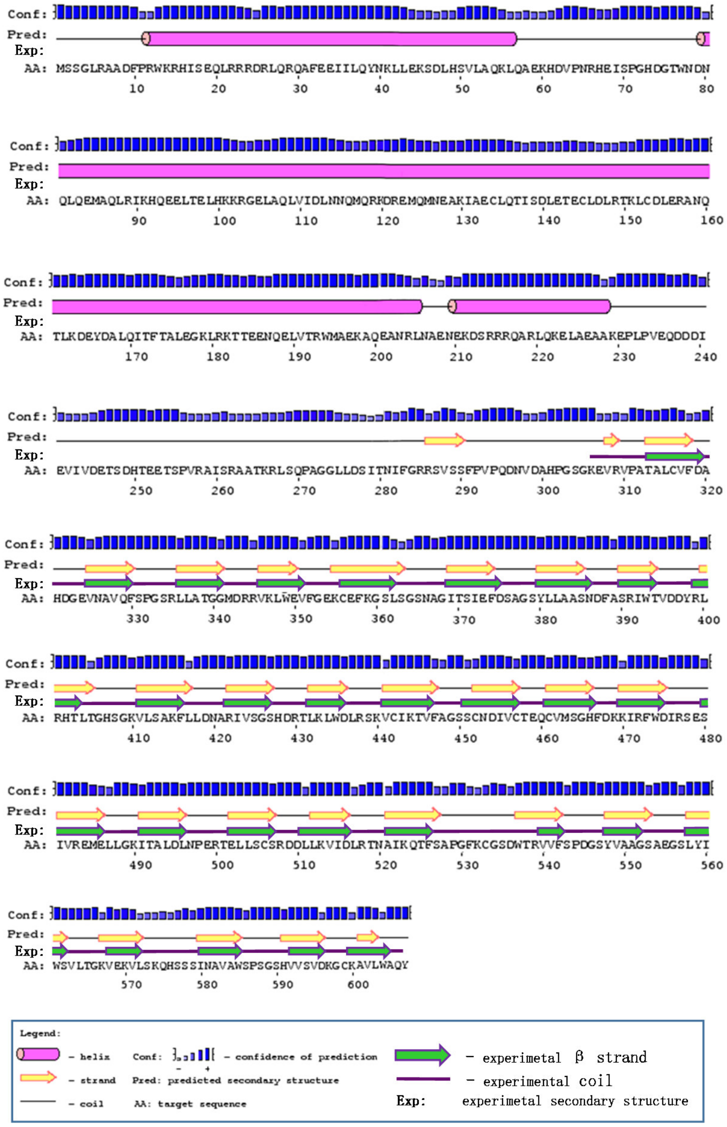


Figure 66: Human ATG16L1 with experimental structural elements for the WD-40 domain compared with the PSIRPED v3.3 output (predicted secondary structure). Purple line represents loops and turns in the structural model. Green arrows with purple outline represent β-strands in the structural model.

Construct which resulted crystals started at position G303. However there was no electron density for residues 303-306 which suggests these residues are flexible with high degree of freedom. Glutamate (E) at position 307 is the first residue with defined electron density for the α carbon. According to the secondary structure prediction, the first β-strand starts with R286 to S290 and the second β-sheet is V309-R310 (Figure 66A). The fact that R286-Y607 construct never resulted in a crystal suggests the peptide stretch from R286-E307 is a flexible tail which might prevent crystal contacts from forming or from persevering long enough for crystal growth. ATG16L1 E307-Y607 arranges into a complete WD-40 domain, which supports the notion that the predicted R286-S290 β-strand is an algorithmic artifact, or at the very least, that it is not part of the WD-40 domain. The dipeptide V309/R310, predicted to be a short β-strand, is part of the N-terminal tail of the G303-P607 structure. Thus, this short strand is clearly an algorithmic artifact. These first two predicted β-strands are thus both incorrectly predicted as belonging to the WD-40 propeller. However, the majority of the predicted β-strands are quite accurate within one to two residues error margin, with the exception of last β-strand and the seventh β-strand from the C-terminus (Figure 66). The seventh β-strand was predicted from W537 until F542. Structure of WD-40 of ATG16L1 revealed this strand is much shorter spanning from V540 to F542. The last β-strand was predicted from A601 until L603, but in reality is four residues longer, from K600 to A605.

5.2.4 Discrepancies between wild type and crystalized WD-40 domain

Crystals are periodic arrangements of fundamental building blocks (proteins), with a very limited number of interactions forming and maintaining the network of proteins together in a crystal. Inter-molecular interactions are not only few but also weak, and very sensitive to environmental change. The few, weak, essential interactions are inherent of the particular protein and cannot be predicted beforehand. Thus, manipulation of the protein is essential in to overcome this obstacle and increase the chances of crystal contact formation. For this purpose, charged, flexible residues with high surface entropy can be successfully mutated to alanine for improved crystallization. In the case of WD-40 propeller of ATG16L1, surface entropy reduction mutations (E354A/K355A) were necessary but not sufficient for crystallization. Surface entropy reduction mutations, shown in Figure 67 on the gray molecule, together with solvent exposed side chains of the neighboring molecule show E354A and K355A reduced long flexible side chains and eliminated steric hindrance, which would have otherwise prevented the two molecules from coming close together.

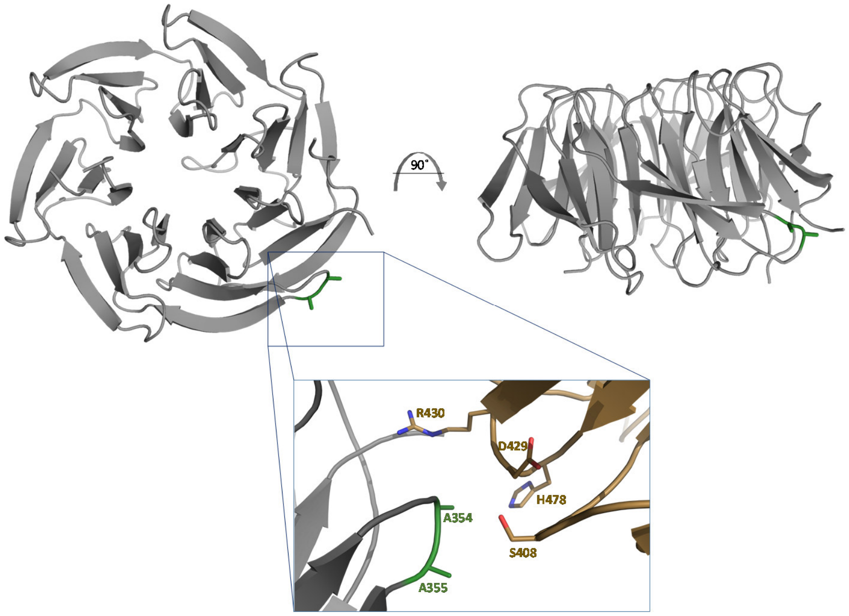


Figure 67: Surface entropy reduction mutations on WD-40 domain of ATG16L1. In green are two alanine residues, which in wild type are glutamine and lysine (E354A/K355A). Sand colored molecule is symmetry-related, neighboring protein and solvent exposed side chains.

In addition to the two surface entropy mutations (E354A/K355A), carboxy-terminal tyrosine is mutated to a proline (Y607P). Tyrosine is present in the human ATG16L1 while proline is in the mouse ATG16L1 (Figure 17). This mutation should not lead to any functional disruption since mouse and human ATG16L1 have high sequence identity (95%). The carboxyl terminus is positioned in such a way that it forms hydrogen bonds with the backbone on the neighboring molecule (Figure 68A). C-terminal proline might form Van der Waals interaction with the phenylalanine from the neighboring, symmetry-related molecule in the crystal (F467) (Figure 68B). If the human wild type tyrosine is modeled in place of proline, it becomes evident the same crystal contacts could not be maintained due to steric clashes (Figure 69). However, a single C-terminal tyrosine to proline substitution was not enough to produce crystals. Thus, it was a combination of both surface entropy reduction mutants and the C-terminal tyrosine to proline substitution resulted in elimination of steric obstacles and allowed formation of the few, weak crystal contacts sufficient for successful crystal growth (Figure 68B).

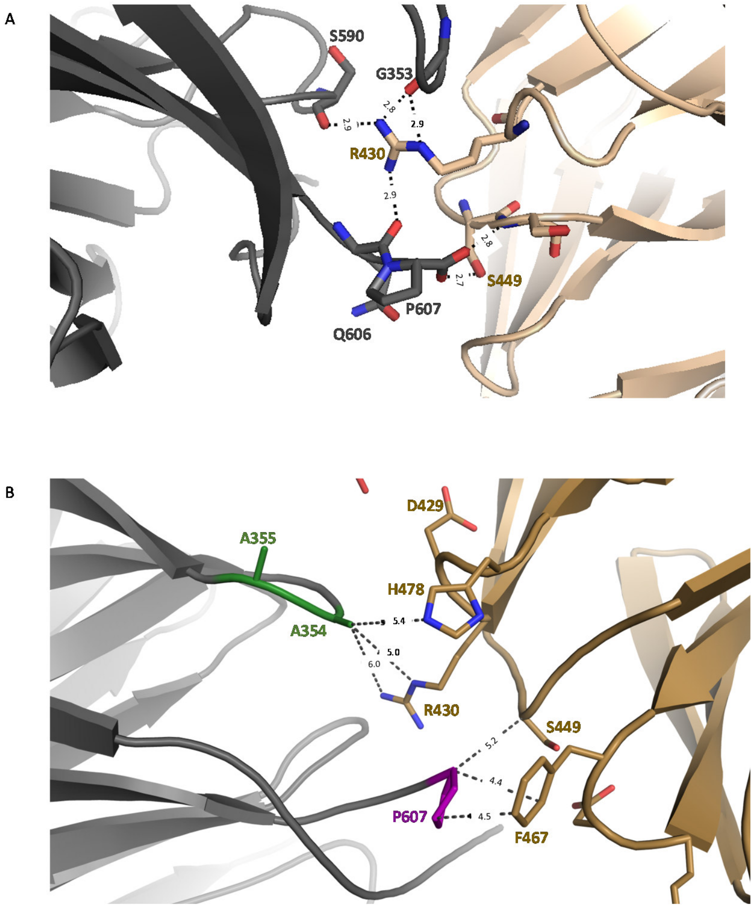


Figure 68: A) Hydrogen bonds between two proteins in the crystal. Proline at the carboxyl terminal of WD-40 propeller of ATG16L1 is labeled on the gray molecule and distances are shown, with dashed lines (in Å) between two symmetry related molecules (gray and sand). B) Wan der Waals contacts between two proteins in the crystal. Gray molecule has A354 and A355 (green) and P607 (purple) labeled. Distances to the side chains on the neighboring molecule (sand) are shown with a dashed line, in Å.

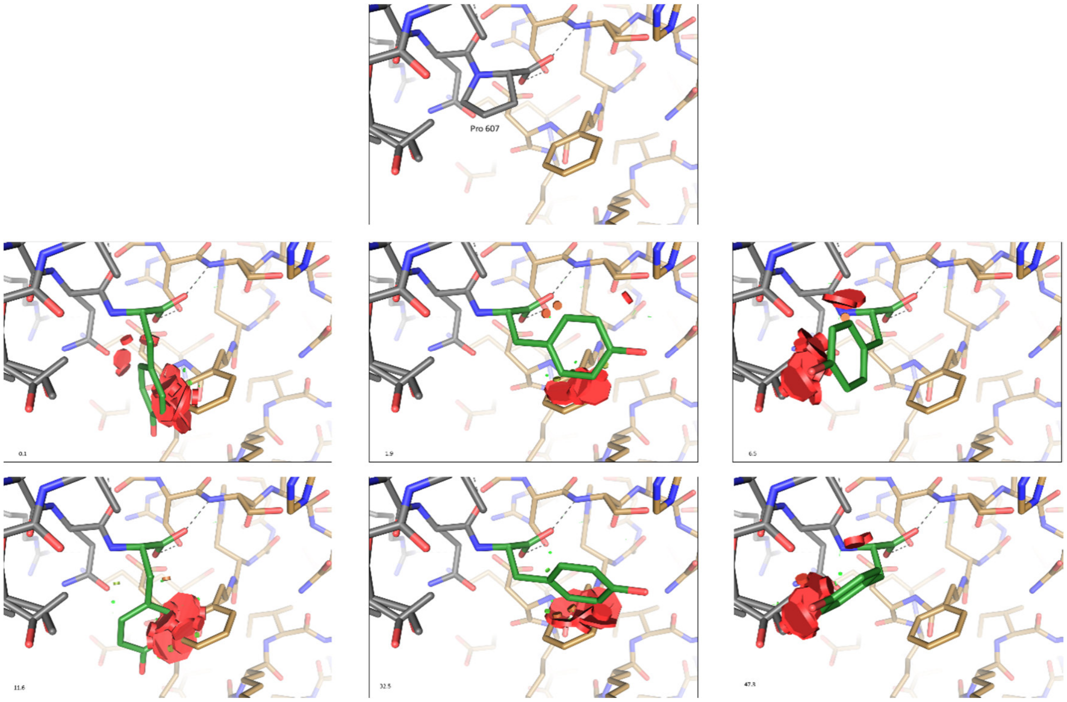


Figure 69: Model of tyrosine at the C-terminal end in crystal structure of WD-40 of ATG16L1. The top panel shows proline at C-terminus. Six panels represent modeled tyrosine (green) at the C-terminus. Red circles represent steric clashes. Each of six panels represents one rotamer of tyrosine, with its probability value in the lower left corner of the panel. Two neighboring molecules are shown in gray and sand color.

5.2.5 Quaternary structure of WD-40 propeller of human ATG16L1

Human ATG16L1 is a member of core autophagy machinery, involved in initiation and elongation step. It is a 607 amino acid long protein, roughly composed of three main functional units: N-terminal ATG5 binding domain (ATG5 BD), coiled coil domain (CCD), and C-terminal WD-40 propeller. Recently, high resolution structure of C-terminal WD-40 domain was solved during the course of this work (Figure 54). C-terminal domain of human ATG16L1 folds into a typical, seven bladed WD-40 propeller. Up to date, both ATG5 BD and CCD of yeast ATG16 have been solved (Figure 70). In the case of human ATG16L1, Otomo *et al.* recently solved the structure of the N-terminal ATG5 BD, a short helix involved in interaction with ATG5 (Mizushima *et al.* 1999; Otomo *et al.* 2013). Structural studies of the yeast ATG16 (Fujioka *et al.* 2010) together with biophysical analysis of CCD of human ATG16L1 (Parkhouse *et al.* 2013) show that the CCD assembles into a parallel dimer, resulting in dimerization of ATG16L1, and in turn dimerization of the ATG12~ATG5/ATG16 complex. By analogy, this would indicate that the WD-40 propeller of ATG16L1 may also form a WD-40/WD-40 interfacing dimer.

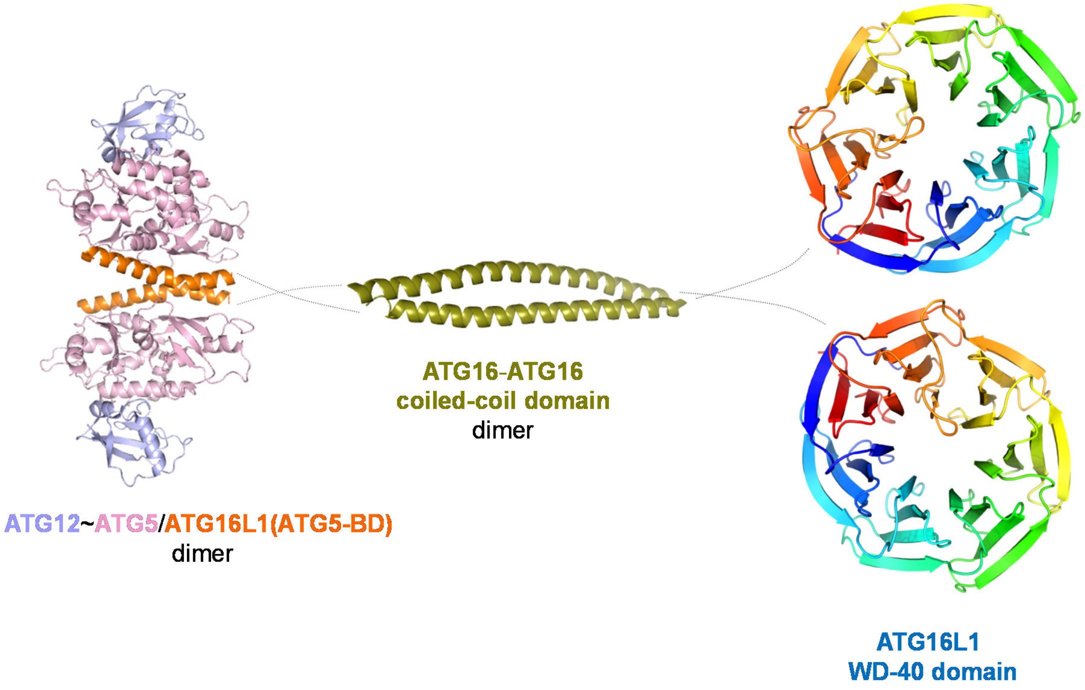


Figure 70: Current state of structural knowledge of human and yeast ATG16. Structure of human ATG16L1 ATG5 binding domain (in orange) has been solved in complex with ATG5~ATG12 conjugate (ATG5 in purple, ATG12 in light blue) (Otomo *et al.* 2013)*. The structure is here modeled into a dimer using PyMOL. Structure of the coiled-coil domain of yeast ATG16 (in green) was solved and showed that the CCD domain assembles into a parallel dimer* (Fujioka *et al.* 2010) *indicating that ATG16 self-dimerizes. WD-40 domain is solved in this work and is represented here as rainbow colored from N to C-terminus.*

In biophysical studies, WD-40 propeller of ATG16L1 behaves as a monomer (Figure 36, Figure 39). In the crystal, two symmetry mates of the propeller are present with the N-terminus pointing in the same direction (Figure 71). Such orientation of WD-40 propellers is in agreement with the physiological parallel dimer of the full length ATG16L1. AREAIMOL (Saff & Kuijlaars 1997; Lee & Richards 1971) calculated the buried surface area of the interface of 3367Å^2 (Table 36). Contacts between the two molecules are dominated by hydrophobic interactions. It is possible this interaction is a result of high protein concentration and crystallographic forces.

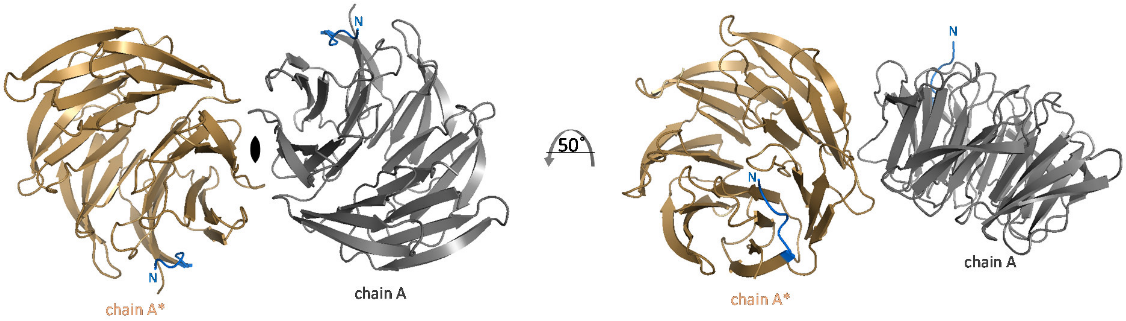


Figure 71: Symmetry mates of WD-40 domain of ATG16L1. Two symmetry related molecules of the WD-40 domain of human ATG16L1 colored in gray (chain A) and sand (chain A). N-terminal tail is labeled in blue on both molecules.*

Total area of chain A	11499.0 Å^2
Total area of chain B	11494.4 Å^2
Total overall area	22993.4 Å^2
Total contact area of chain A	3367.5 Å^2
Total contact area of chain B	3366.1 Å^2
Total overall contact area	6733.6 Å^2

Table 36: AREAIMOL calculated surface area of two symmetry related WD-40 molecules

Since the WD-40 domain of ATG16L1 is a monomer is solution, it might form a dimer only when certain forces are applied: either crystallographic forces or dimerization of the entire ATG16L1 molecule. The dimerization of ATG16L1 through the coiled coil domain might bring the two WD-40 domains in the correct orientation and close enough to facilitate domain dimerization as well. Alternatively, the overall dimer conformation of the ATG16L1 might be enough to maintain the WD-40/WD-40 platform with no additional dimerization surfaces on the propeller. However, lack of full length structure of ATG16L1 prevents us from knowing if the WD-40 domains really generates an even greater interaction platform, or if they have enough freedom of movement within the dimer to act independently from one another. Structural studies of full length ATG16L1 are needed to answer this question.

5.2.6 *B*-factor analysis and DWD/WDxR box

Proteins are not static molecules. Peptide backbone and atoms in the side chains change due to kinetic energy and thermal motion. The *B*-factor of a protein structure reflects displacement of atoms from their average position. Low *B*-factors usually indicate rigid parts of the protein. High *B*-factors indicate areas of the protein with high flexibility and mobility. Based on the *B*-factor map of the WD-40 domain of ATG16L1 (Figure 72), the constriction of the propeller has low *B*-factors, implying a rigid and stable fold with low flexibility. The constriction is made up of strands A of each blade. *B*-factor increases towards the outside of the WD-40 propeller, with strands A having the lowest and strands D having the highest *B*-factor. As far as β-strands are concerned, strands D belonging to blades 2, 3, 4 exhibit the highest level of flexibility. Blade 3 is particularly interesting for two reasons: strand C of blade 3 has highest *B*-factors of all strands C, and blade 3 loops facing the bottom of the propeller (between strands A and B and strands C and D) are also quite flexible. Region between strand C of one blade and strand A of the next (including strand D and the loops) is where greatest variation is found in WD-40 propellers (Figure 59). The second variable region is the loop between strands A and B. In the case of WD-40 propeller of ATG16L1, the loop between A and B could not be completed during model building due to the lack of electron density in this region. Absence of electron density in the part of the protein could be due to one of two reasons: either there was a cleavage event, or the disorder level is too high. However, since dissolved crystals did not result in a double band nor in a size decrease when dissolved and loaded on an SDS PAGE gel (Figure 47), high level of disorder in this loop is a more likely explanation for the absence of electron density.

Notably, the loop between strand A and B of the blade 3 participates in formation of no contacts with neighboring symmetry related mates, which might be a reason for the lack of electron density in that region. Alternatively, the flexibility of this loop might have functional relevance.

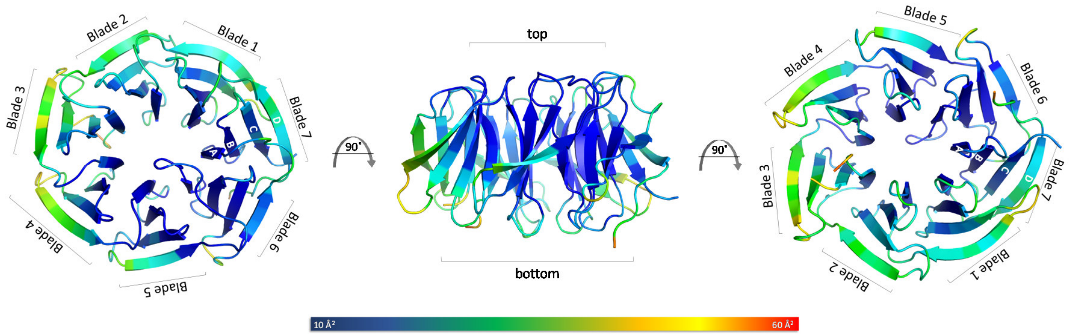


Figure 72: B-factor distribution on WD-40 propeller of ATG16L1. Core of the propeller has low B-factors, indicating low flexibility of the side chains. B-factors increase towards the outside of the propeller, with highest B-factors on blade 3 of the propeller.

Based on the *B*-factor analysis, blade 3 of WD-40 propeller of ATG16L1 seems to be the most flexible part of the protein in the crystal. As alluded earlier, this flexibility might either be functionally important, or arise due to the lack of stabilizing crystal contacts between this area of the molecule with neighboring mates. Closer analysis at the crystal contact distribution (Figure 73) indicates no obvious or drastic reduction between contacts formed by blade 3 compared to the contacts formed by other blades of the propeller. This observation allows the conclusion that the flexibility of blade 3 has functional implications. However, further studies are needed to confirm this assumption.

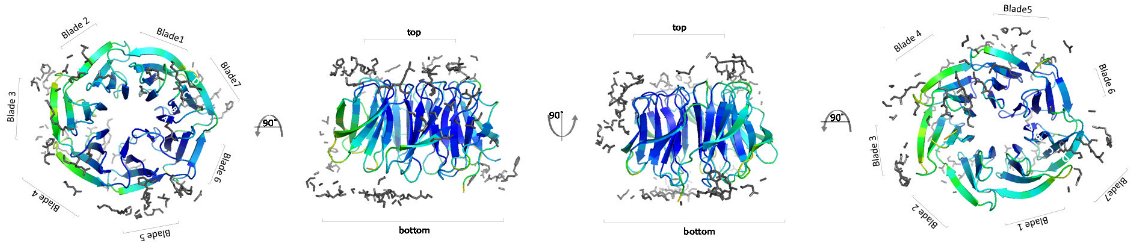


Figure 73: Crystal contacts formed by WD-40 propeller of ATG16L1 in a crystal. WD-40 propeller is represented in colors according to B-factors .Crystal contacts are depicted in gray, in the radius of 6 Å.

The top, narrow end of the WD-40 of ATG16L1 exhibits lower *B*-factor distribution compared to the bottom of the propeller. Thus, based on the *B*-factor analysis, it could be assumed that the functional surface of the WD-40 propeller of ATG16L1 is either on the bottom of the propeller, and/or at the circumference in the region of blade 3.

Support for the hypothesis that blade 3 of ATG16L1 WD-40 propeller has a functional relevance is also found in the literature. He *et al.* defined the DDB1-binding and WD-40 repeat (DWD) box, a 16 amino acid stretch present on DCAFs. Proteins containing DWD box are not limited to one, but can sometimes contain two DWD boxes, however seldom three. He *et al.* showed that ATG16L1 binds DDB1 via this motif (He *et al.* 2006). DWD box identified by He *et al.* is located on the strand B and strand C (including the loop between them and part of the loop between strands C and D) of blade 3 of the ATG16L1 WD-40 domain (Figure 74).

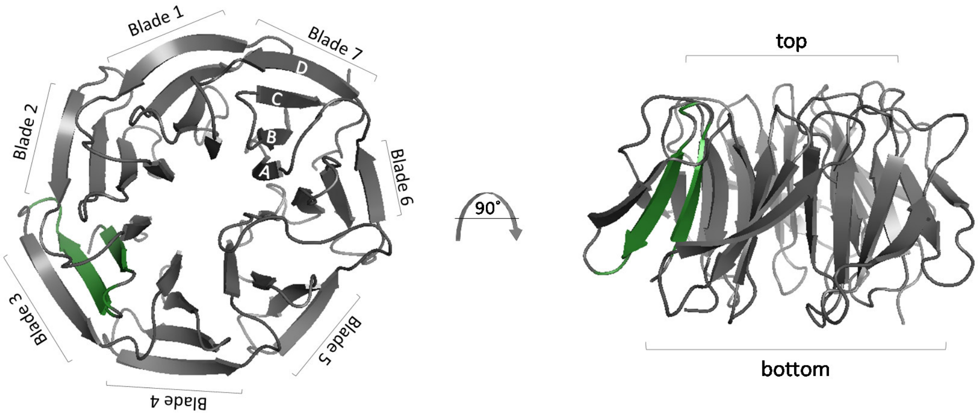


Figure 74: DWD (the DDB1-binding and WD-40 repeat) box (green) defined by He et al. on the WD-40 propeller of ATG16L1 (gray).

Point mutation of an arginine to alanine at position 424 on ATG16L1 disrupted the binding with DDB1 (He *et al.* 2006). However, in the DWD box identified by He *et al.* the only arginine is at position 438. For the further discussion purposes, the reported arginine 424 and arginine 438 are treated as one and the same. From the structural model of WD-40 of ATG16L1, this proposed functionally essential residue is in the loop between strands C and D of blade 3, facing the bottom of the propeller (Figure 75).

DWD box is an extension of a more general motif found on a variety of WD-40 propellers. This conserved DxR or WDxR motif is commonly found on WD-40 substrates of DDB1/CUL4. Similarly to the DWD box, WDxR motif often occur twice on the propeller, at the end of strand C of two consecutive blades (Angers *et al.* 2006).

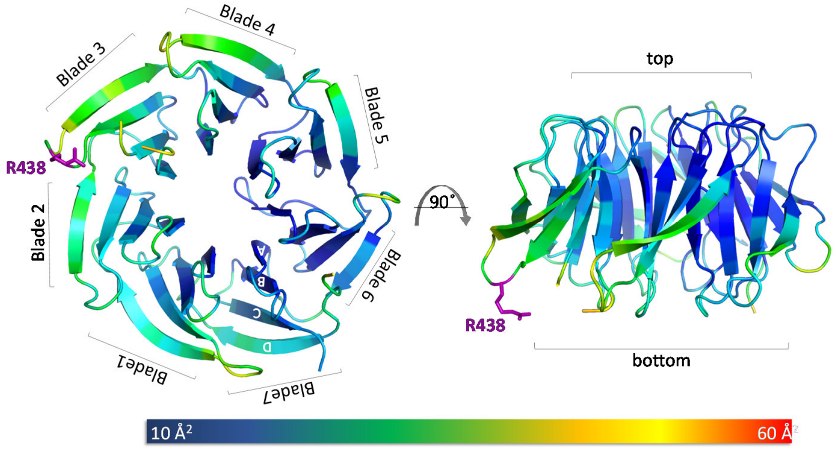


Figure 75: Position of conserved arginine involved in DDB1 binding on WD-40 domain of ATG16L1. In this figure, bottom and side view of B-factor distribution of WD-40 domain of ATG16L1 is presented. Arginine (R438 labeled in purple) is proposed to be the conserved residue on the DWD box present on DCAFs binding to DDB1-CUL4 (He *et al.*, 2006)*. On WD-40 domain of ATG16L1 this arginine is on the flexible loop between strands C and D on blade 3.*

In the case of ATG16L1, one WDxR motif, WDLR labeled in red in Figure 76, is contained within the DWD box, on the loop between strands C and D of blade 3. The second WDxR motif, labeled in blue in Figure 76, is on the corresponding loop on the neighboring blade 4. Both motives are located on the bottom surface of the propeller.

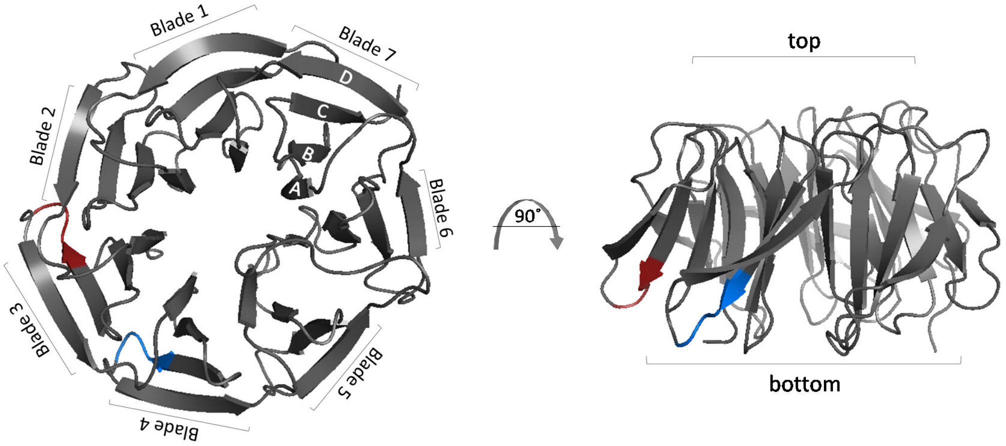


Figure 76: WDxR motives on the WD-40 of ATG16L1. WD-40 propeller is labeled in gray. Two WDxR motives are labeled according to the following scheme: WDLR in red and WDIR in blue.

Based on the high *B*-factors present identified in the crystal structure of WD-40 of ATG16L1 and the functional assay performed by He *et al.*, blade 3 of the WD-40 of ATG16L1 with its loops facing the bottom of the propeller could play a major functional role of this domain. However, the functional role of the WDxR containing region is not necessarily part of the interaction surface between DCAFs and DDB1/CUL4, as assumed by He *et al.*. In case of DDB2, the WDxR motif is not part of the surface involved in the binding to DDB1 or to DNA. Nevertheless, disruption of the WDxR motif on DDB2 leads to a nonfunctional protein, indicating that intact WDxR motif is important for proper folding and/or maintains the stability of the WD-40 propeller rather than being essential for interaction. (Scrima *et al.* 2011)

It is evident that circumference of the WD-40 domain of ATG16L1 does not exhibit the same level of flexibility all around. D strands of blades 5, 6, 7 and 1 seem to be less flexible compared to D strands of blades 2, 3 and 4. Since this is not an artifact of crystal contact distribution (Figure 73), it might indicate that, if dimerization of the WD-40 domain of ATG16L1 takes place, the surface facilitating the dimerization is composed of a part of a region defined by blades 5, 6, 7 and 1. In this case, it makes sense to keep the stability of this half of the propeller. Indeed, in the two symmetry mates which bury enough surface area to have physiological relevance (Figure 71, Table 36), the interface region is composed mainly of blades 5 and 6 of each propeller. On the other hand, it would also mean that part of the propeller remains unused. This would only be justified if by dimerization, the two bottom surfaces (with parts of blade 3) create a greater super-interaction surface, compensating for the part used in dimerization. Additional analysis and identification of the WD-40 domain surfaces responsible for interaction with other macromolecules might provide further clues on whether and how the WD-40 domain of ATG16L1 interact with each other.

5.2.7 Surface analysis

WD-40 propellers provide three main surfaces for interaction with other macromolecules: the narrow top, the wider bottom and the circumference. Analysis of protein surface in terms of electrostatic potential shows the top of the WD-40 domain of ATG16L1 has a positive charged area on one half around the constriction (Figure 77A, left image). The circumference of the WD-40 domain of ATG16L1 is mainly neutral (Figure 77A, two middle images), possibly because the two domains exist relatively closely to each other in the ATG16L1 dimer, and should not repel each other by having two strong charges facing each other. The bottom of the WD-40 domain of ATG16L1 (Figure 77A, right image) shows a strong overall positive electrostatic potential. Based on all three binding surfaces available on the WD-40 domain of ATG16L1, the wide bottom surface seems like the most obvious choice in terms of areas involved in polar interactions.

Analysis of the hydrophobic distribution on the surface of the WD-40 propeller of ATG16L1 provide little further insight into the possible function of this domain. From Figure 77B, the hydrophobic patches are distributed over all three interaction surfaces. Lower half of the circumference area between blades 3 and 4 (Figure 77B, second image from the right) provides the largest hydrophobic surface, indicating the hydrophobic interactions are facilitated through this interface. However, without experimental and structural data of ATG16L1 WD-40 in complex with its interaction partners, other surfaces on the WD-40 domain of ATG16L1 cannot be excluded at this point.

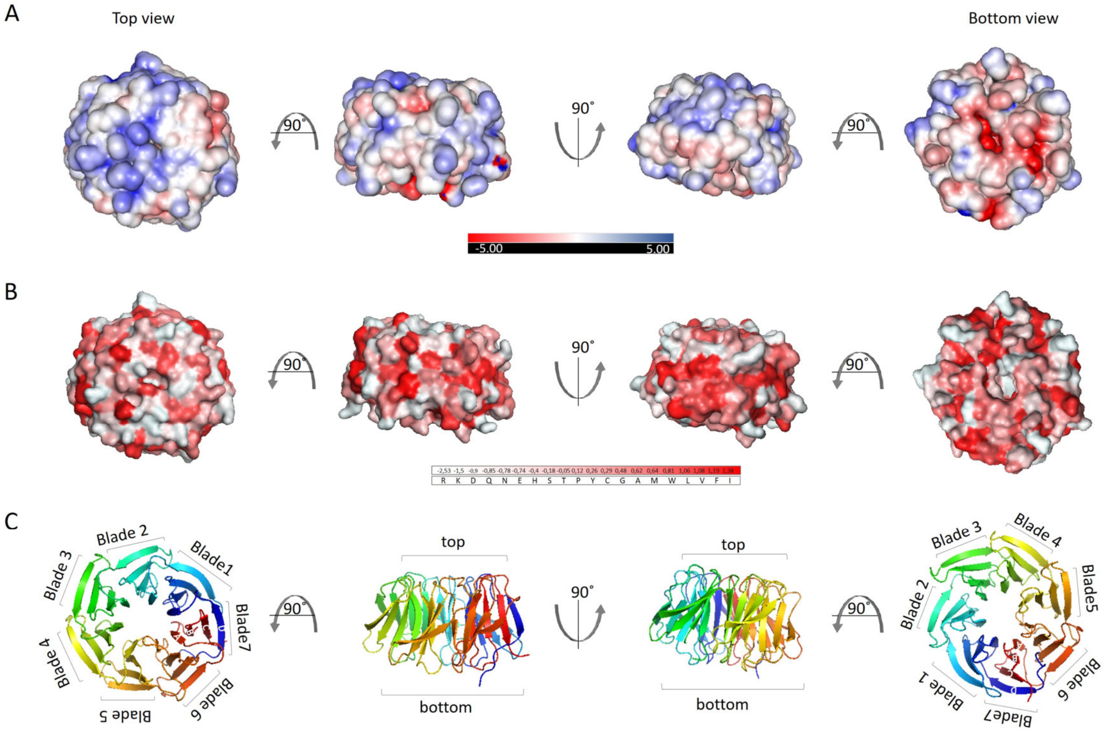


Figure 77: Surface analysis of the WD-40 propeller of human ATG16L1. A) Electrostatic surface potential calculated by ABPS (Baker *et al.* 2001) *and mapped onto the surface of WD-40 propeller of ATG16L1 in PyMOL. Areas marked in red correspond to the negative electrostatic surface potential; areas marked in blue correspond to the positive electrostatic surface potential. The surface potential is contoured from -5 to 5* k_BT/e *(*k_B *is the Boltzmann constant, T is temperature in Kelvin, e is charge on the electron). B) WD-40 domain of ATG16L1 colored by hydrophobicity of the residues according to the Eisenberg scale* (Eisenberg *et al.* 1984)*. C) The orientation of the propeller corresponding to the surface shown directly above it.*

5.3 Polymorphism of human ATG16L1

A naturally occurring missense single nucleotide polymorphism of human ATG16L1, resulting in T300A substitution, has been identified as one of risk alleles contributing to Crohn's disease. Crohn's disease is a chronic inflammatory disease that can affect any part of the gastro-intestinal track. Experiments using the T300A knock-in mice revealed a reduced level of bulk and selective autophagy in multiple cell types and diminished clearance of pathogens by macrophages (Lassen *et al.* 2014; Murthy *et al.* 2014). Functional studies have further shown that T300A ATG16L1 is more prone to cleavage by caspase 3 and caspase 7 (Lassen *et al.* 2014; Murthy *et al.* 2014). Biophysical studies of full length ATG16L1 identified a degradation product at position 307 (Figure 21). Taken together, this evidence presents a possibility that the area between 300 and 307 might be inherently fragile and the SNP drastically increases its susceptibility for caspase cleavage.

Cleavage of ATG16L1 T300A full length leads to overall decrease of ATG16L1 protein levels, but does not result in the full loss-of-function of autophagy. This is consistent with current understanding that Crohn's disease is a complex disorder caused by a combination of genetic and environmental factors. Additionally, this observation gives rise to the notion that the cleaved ATG16L1 might still remain partially functional and sufficient to carry out basal autophagy.

When cleavage occurs at position 300, intact WD-40 domain (307-607) is separated from the full length protein and both fragments are still present after 4 hr of caspase activation (Lassen *et al.* 2014). Thus, the N-terminal half of ATG16L1 is intact after caspase cleavage and sufficient to coordinate bulk autophagy process. On the other hand, WD-40 domain serves as a recruiter or a linker between autophagy and other cellular process and has no functional contribution to the basal autophagy. Thus, caspase cleavage event detaches the WD-40 domain from the rest of the autophagy machinery, which may still retain its ability to form a complex with its interaction partners but can no longer act as a physical linker between pathogen recognition and autophagy.

In addition to the natural SNP, five splicing variants have been described for human ATG16L1 (UniProt identifiers: Q676U5-1, Q676U5-2, Q676U5-3, Q676U5-4 and Q676U5-5) occurring in tissue-specific manner. A novel isoform of mammalian ATG16L1 has been identified and termed ATG16L2 (UniProt: Q8NAA4) (Ishibashi *et al.* 2011). Sequence comparison between human ATG16L1 and ATG16L2 (Figure 78) indicates overall sequence identity of 38.8 % and 67.2 % similarity.

```
ATG16L1      -MSSGLRAADFPRWKRHISEQLRRRDRLQRQAFEEIILQYNKLLEKSDLHSVLAQKLQAEKHDVPNRHEISPGHDGTWNDN-----QLQEMAQLRIKHQEELTELHKKRGELAQLVIDLN      114
ATG16L2      MAGPGVPGAPAARWKRHIVRQLRLRDRTQKALFLELVPAYNHLLEKAELLDKFSKKLQPEPNSVTPT-----THQGPWEESELDSDQVPSLVALRVKWQEEEEGLRLVCGEMAYQVVEKG      115
               . *: .*   ****** .*** *** *:  * *::  **:****::* . :::*** * ..*         *:* *::.     *: .:. **:* ***   *:   **:*  *::

ATG16L1      NQMQRKDREMQMNEAKIAECLQTISDLETECLDLRTKLCDLERANQTLKDEYDALQITFTALEGKLRKTTEENQELVTRWMAEKAQEANRLNAENEKDSRRRQARLQKELAEAAKEPLPV      234
ATG16L2      AALGTLESELQQRQSRLAALEARVAQLREARAQQAQQVEEWRAQNAVQRAAYEALRAHVGLREAALRRLQEEARDLLERLVQRKARAAAERNLRNERRERAKQARVSQELKKAAKRTVSI      235
               :   : *:* .::::*     :::*.    :   :: : .  * . :  *:**:  .   *. **:  ** ::*: * : .**: * . * .**: .* :***:.:** :***. : :

ATG16L1      EQDDDIEV-IVDET---SDHTEETSPVRAISRAATKRLSQPA----------GGLLDS---ITNIFGRRSVSSFPVPQDNVDTHPGSGKEVRVPATALCVFDAHDGEVNAVQFSPGSRLL      337
ATG16L2      SEGPDTLGDGMRERRETLALAPEPEPLE---KEACEKWKRPFRSASATSLTLSHCVDVVKGLLDFKKRRGHSIGGAPEQRYQIIP-VCVAARLPTRAQDVLDAHLSEVNAVRFGPNSSLL      351
             .:  *     : *       : * .*:.   : * :: .:*           .  :*    : ::  **. *   .*::. :  *     .*:*: *  *:*** .*****:*.* * **

ATG16L1      ATGGMDRRVKLWEVFGEKCEFKGSLSGSNAGITSIEFDSAGSYLLAASNDFASRIWTVDDYRLRHTLTGHSGKVLSAKFLLDNARIVSGSHDRTLKLWDLRSKVCIKTVFAGSSCNDIVC      457
ATG16L2      ATGGADRLIHLWNVVGSRLEANQTLEGAGGSITSVDFDPSGYQVLAATYNQAAQLWKVGEAQSKETLSGHKDKVTAAKFKLTRHQAVTGSRDRTVKEWDLGRAYCSRTINVLSYCNDVVC      471
             **** ** ::**:*.*.: * : :*.*: ..***::** :*  :***: : *:::*.* : : :.**:**. ** :*** * . : *:**:***:* ***    * :*: . * ***:**

ATG16L1      TEQCVMSGHFDKKIRFWDIRSESIVREMELLGKITALDLNPERTELLSCSRDDLLKVIDLRTNAIKQTFSAPGFKCGSDWTRVVFSPDGSYVAAGSAEGSLYIWSVLTGKVEKVLSKQHS      577
ATG16L2      GDHIIISGHNDQKIRFWDSRGPHCTQVIPVQGRVTSLSLSHDQLHLLSCSRDNTLKVIDLRVSNIRQVFRADGFKCGSDWTKAVFSPDRSYALAGSCDGALYIWDVDTGKLESRLQGPHC      591
              :: ::*** *:****** *.   .: : : *::*:*.*. :: .*******: *******.. *:*.* * *********:.***** **. ***.:*:****.* ***:*. *.  *.

ATG16L1      SSINAVAWSPSGSHVVSVDKGCKAVLWAQY      607
ATG16L2      AAVNAVAWCYSGSHMVSVDQGRKVVLWQ--      619
             :::*****. ****:****:* *.***
```

*Figure 78: Sequence alignment between human ATG16L1 and human ATG16L2 by Clustal Omega. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties.*

Both isoforms are similar in domain arrangement, in that they both have an N-terminal ATG5 binding domain, middle coiled-coil region and a C-terminal WD-40 domain. Consequently, both isoforms of mammalian ATG16L1 have been shown to form dimers (both ATG16L2/ATG16L2 and hetero-dimers ATG16L1/ATG16L2) and bind ATG5~ATG12 conjugate (Ishibashi *et al.* 2011). However, ATG16L2 isoform was not found on the growing autophagosomal membrane (Ishibashi *et al.* 2011), indicating this isoform is dispensable for bulk, canonical autophagy and cannot compensate for the loss of function of ATG16L1. However, in an independent study, knock-down of ATG16L2 resulted in higher occurrence of autophagy (Lipinski *et al.* 2010) suggesting that ATG16L2 is indeed involved in canonical autophagy as a negative regulator.

Regardless of the involvement of ATG16L2 in canonical autophagy, sequence analysis of the C-terminal domain (303-607) between ATG16L1 and ATG16L2 indicates 51.2% sequence identity and 78.1 % similarity. Due to high similarity between the C-terminal domain of the two isoforms, ATG16L2 might be important for other types of autophagy, such as xenophagy. Further work is needed to clearly define if and which role ATG16L2 pays in which type of autophagy.

5.4 Putative interaction partners of WD-40 domain of ATG16L1

Cellular pathways are often regulated by transient assemblies of dynamic complexes. Common feature of all WD-40 domains characterized so far is a scaffolding function. As a scaffold, WD-40 domains serve as an interaction platform for proteins, peptides and nucleic acids. Other than being essential for autophagy, ATG16L1 has been implicated in pathogenic control (Kuballa *et al.* 2008), antigen presentation (Cooney *et al.* 2010) and Crohn's disease (Barrett *et al.* 2008). More precisely, the WD-40 domain of ATG16L1 has been shown to bind DDB1 (He *et al.* 2006), CARD domain of NOD1 and CARD1 domain of NOD2 (Ramjeet *et al.* 2010; Travassos *et al.* 2010), TIR domain of TLR2 and a short cytosolic peptide of TMEM59 (Boada-Romero *et al.* 2013). Boada-Romero *et al.* have identified an ATG16-binding motif ([YW]-$X_{(2,6)}$-[ED]-$X_{(2,6)}$-[YWF]-X_2-L) in CARD1 domain of NOD2, TIR domain of TLR2 and in TMEM59, responsible for an interaction with WD-40 domain of ATG16L1. This novel binding motif is incomplete on CARD domain of NOD1, and there is conflicting evidence if NOD1 interacts with ATG16L1 (Boada-Romero *et al.* 2013; Travassos *et al.* 2010). Proposed ATG16L1-binding motif is highlighted in red on each molecule in Figure 79, and is surface exposed and in principle accessible for interaction by the WD-40 domain of ATG16L1. However, no corresponding binding motif has been identified on the side of the ATG16L1.

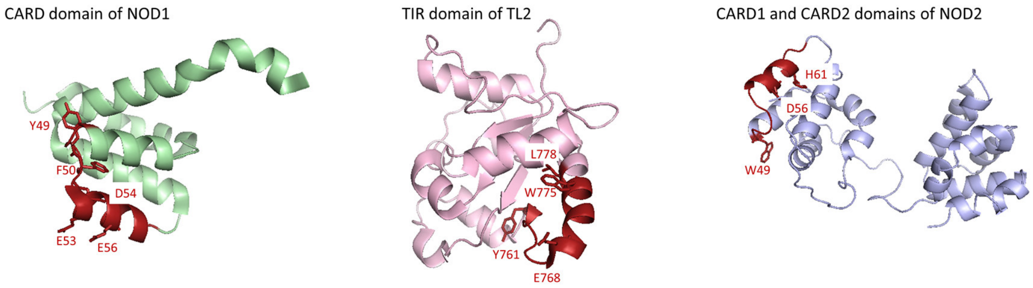


Figure 79: ATG16L1-binding motif on the proposed interaction partners. CARD domain of NOD1 is represented as cartoon in green (PDB code: 2NZ7, (Srimathi *et al.* 2008)*). TIR domain of TLR2 is represented as cartoon in pink (PDB code: 1FYW, (Xu et al. 2000)). CARD1 and CARD2 domains of NOD2 are represented as cartoon in blue, and modeled using Phyre2 server (93% modelled at >90% confidence). ATG16L1-binding motif suggested by Boada-Romero et al. is highlighted in red, with residues proposed to be essential for interaction individually labeled.*

Since WD-40 domain of ATG16L1 is dispensable for bulk autophagy, it is only logical that this domain acts as a bridge between autophagy and other cellular processes (Figure 2). During the course of this work, different techniques have been used to test and characterize these linking interactions. Pull down assays (Figure 40) using recombinantly expressed and purified domains of proteins showed no binding between WD-40 domain of ATG16L1 and its suggested interaction partners. A more sensitive method (Figure 43) showed the same lack of interaction. In the experimental setting used in these experiments with recombinantly purified proteins, the interaction between WD-40 domain of ATG16L1 with proposed interaction partners is not detectable. Based on this information it is safe to assume that WD-40 domain of ATG16L1 does not directly bind suggested proteins, but at least one more factor mediates the interaction. This (or these) additional, so far undefined mediator(s) would naturally not be present in purified samples used for pull downs using samples from heterologous expression hosts and MST.

There are, of course, multiple disadvantages when studying interactions in vitro. Since it is not a true representation of the environment under which cellular interactions take place, some important factors might be missing as well as post translational modifications. Even though phosphomimetic mutants were created on predicted sites, these might not be the only sites or the artificial modification is not fully compatible with the natural modification. In addition, some co-factors or small molecules might be needed for facilitate the complex formation. Alternatively, additional protein (or more than one) might be mediating or stabilizing the complex formation, which remained unidentified so far. Towards the end of this work, human embryonic kidney cells (HEK 293E) were transiently transfected with full length ATG16L1, WD-40 domain of ATG16L1 and N-terminal half of the ATG16L1 in the search for mediators. Samples were analyzed using mass spectrometry and HPLC, hoping to confirm already mentioned proteins as interaction partners, or identify novel ones. Following few rounds of optimization possible during the course of this work no definite conclusions were possible. Further optimization is needed to improve transfection protocol of HEK 293E with multiple plasmids and establish a more definite time course of autophagy in this cell line, using the cultivation conditions in our lab. In addition to autophagy being a very dynamic process, it is still unclear at which point exactly it interferes with pathogen detection. It seems un-likely that NOD2 and all downstream events remain stimulated for prolonged periods of time. Thus, additional time course experiments are needed to establish a precise time line of events. According to Travassos *et al.* NOD2 recruits ATG16L1 to the site of bacterial entry, and the two proteins co-localize with invading bacteria. They also report that ATG16L1 lacking the N-terminal ATG5 binding domain also co-localized with NOD2 and invading bacteria (Travassos *et al.* 2010). This puts the spotlight on the WD-40 of ATG16L1. However, NOD2-ATG16L1 relationship is more complicated than a simple protein-protein interaction. Additional reports show that ATG16L1 regulates NOD2 in autophagy independent manner (Sorbara *et al.* 2013) and that an additional autophagy protein, IRGM, links NOD2 with ATG16L1 and the autophagy machinery (Chauhan *et al.* 2015). As more and more evidence regarding NOD2-ATG16L1 interaction and the connection between autophagy and pathogen detection comes to light, it becomes evident how complex and tightly regulated this association is. It cannot be assumed that other proposed interaction partners of WD-40 domain of ATG16L1 have reduced level of complexity and are under less strict spatio-temporal control. More research is needed in to unwind and understand this intricately connected cellular network.

6 Conclusions and Outlook

ATG16L1 is a central player in autophagy, a fundamental catabolic process, conserved throughout eukaryotic domain of life. During this work, human ATG16L1 and more specifically the WD-40 domain of human ATG16L1 have been studied using biophysical, biochemical and structural methods. Expression and purification of full length human ATG16L1 did not yield enough protein for crystallization setups. The biophysical analyses of this protein were also difficult due to its large size and its multi-domain architecture which made it readily prone to degradation. Interestingly, the main site of degradation was exactly at the beginning of the C-terminal WD-40 domain.

The crystal structure of C-terminal domain was solved at high resolution and revealed geometric and molecular detail of this domain. As predicted, the C-terminal domain of human ATG16L1 folds into a WD-40 propeller domain. The WD-40 domain of human ATG16L1 is a typical seven bladed propeller with each blade composed of four anti-parallel β-strands. The number of blades and β-strands contained in the domain differs from predicted secondary structure elements. This illustrates the weakness of prediction algorithms, especially in cases of domains with high functional and structural diversity, such as WD-40 propellers.

Biophysical analysis showed this domain was a monomer in solution. Analysis of crystal contacts revealed a large buried area between two molecules of WD-40 domain, with their N-termini pointing in the same direction. Surface area shared between two molecules indicates physiological relevance, while the orientation of the N-termini is reminiscent of a parallel dimer formed by the coiled-coil domain of full length ATG16L1. However, further experiments are needed to define a functional, oligomeric state of the WD-40 domain.

WD-40 domain is among the most abundant eukaryotic domains with the widest range of interaction partners. WD-40 domains offer three surfaces for interactions: narrow top, wide bottom, and the circumference of the propeller. This partially contributes to the functional variability observed in this domain family and makes functional assignment to these domains difficult. The most common feature of WD-40 domains is to serve as a rigid molecular scaffold for protein-protein interactions. Based on this commonality and functional data already published, WD-40 domain of ATG16L1 might offer an interaction platform for proteins and protein complexes linking autophagy with pathogen recognition, control and clearance.

Analysis of *B*-factor distribution and electrostatic surface potential provided insight into the dynamics of the WD-40 domain of human ATG16L1. Namely, blade 3 of the WD-40 propeller together with the bottom, wider face exhibit higher level of flexibility compared to the rest of the molecule. This information was used to help narrow down which part of the propeller is likely to be functionally relevant. Additional evidence for the notion that the blade 3 is involved in interactions between ATG16L1 and its binding partners comes from the literature. A 16 residue DDB1-binding and WD-40 repeat (DWD) box present on DCAFs (DDB1-CUL4 associated factors), also identified on the WD-40 propeller of ATG16L1 is sited on the blade 3 of the WD-40 propeller of ATG16L1, with a functionally essential arginine facing the bottom of the propeller. The DDB1-CUL4 ubiquitin E3 ligase is involved in cell-cycle progression, replication and DNA damage response. However, not all DCAFs have been identified and characterized, and the WD-40 domain of ATG16L1 has been proposed as one such receptor. However, during the course of this work, the interaction between WD-40 domain of ATG16L1 and DDB1 could not be confirmed by the *in vitro* assays using recombinantly expressed and purified proteins.

Apart from DDB1, WD-40 domain of ATG16L1 has been implicated as a binding partner for CARD domain of NOD1, CARD1 domain of NOD2, TIR domain of TLR2 and TMEM59. Various biophysical methods were tested with recombinantly expressed and purified proteins, however no interaction was detected. This observation leads to a hypothesis that these interactions are not direct, but rather mediated by additional factor or factors. To test this idea, transient protein expression in HEK 293E cells was used to identify novel partners of the WD-40 domain. However, due to the lack of time and expertise in the cellular biology and transfection of mammalian cells with multiple plasmids, we were not able to arrive to any conclusive results. Further optimization of transient transfection protocol, induction of autophagy and other cellular pathways, time course studies and thorough analysis of samples by mass spectrometry might still reveal additional molecular factors involved with the WD-40 domain of ATG16L1. Once these factors are identified, it would be interesting to co-crystallize them in complex with the WD-40 domain of ATG16L1 and observe if the blade 3 and the bottom surface of the propeller play a role. Moreover, novel and confirmed interaction partners would provide correct purpose of this large, but so far not fully functionally characterized domain and reveal an accurate link between autophagy and other cellular processes in a mammalian system.

References

Abbott, D. W., Wilkins, A., Asara, J. M., & Cantley, L. C. (2004). The Crohn's disease protein, NOD2, requires RIP2 in order to induce ubiquitinylation of a novel site on NEMO. *Current Biology : CB*, *14*(24), 2217–27. http://doi.org/10.1016/j.cub.2004.12.032

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., … Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 213–21. http://doi.org/10.1107/S0907444909052925

Andrade, R. M., Wessendarp, M., Gubbels, M.-J., Striepen, B., & Subauste, C. S. (2006). CD40 induces macrophage anti-Toxoplasma gondii activity by triggering autophagy-dependent fusion of pathogen-containing vacuoles and lysosomes. *The Journal of Clinical Investigation*, *116*(9), 2366–77. http://doi.org/10.1172/JCI28796

Angers, S., Li, T., Yi, X., MacCoss, M. J., Moon, R. T., & Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*, *443*(7111), 590–3. http://doi.org/10.1038/nature05175

Bajagic, M. (2015). *Structure and Function of Human Autophagy Protein ATG16L1*. Technischen Universität Carolo-Wilhelmina, Braunschweig.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10037–41. http://doi.org/10.1073/pnas.181342398

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., … Daly, M. J. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genetics*, *40*(8), 955–62. http://doi.org/10.1038/ng.175

Beattie, R. M., Croft, N. M., Fell, J. M., Afzal, N. A., & Heuschkel, R. B. (2006). Inflammatory bowel disease. *Archives of Disease in Childhood*, *91*(5), 426–32. http://doi.org/10.1136/adc.2005.080481

Behrends, C., Sowa, M. E., Gygi, S. P., & Harper, J. W. (2010). Network organization of the human autophagy system. *Nature*, *466*(7302), 68–76. http://doi.org/10.1038/nature09204

Beron, W., Gutierrez, M. G., Rabinovitch, M., & Colombo, M. I. (2002). Coxiella burnetii Localizes in a Rab7-Labeled Compartment with Autophagic Characteristics. *Infection and Immunity*, *70*(10), 5816–5821. http://doi.org/10.1128/IAI.70.10.5816-5821.2002

Birmingham, C. L., Smith, A. C., Bakowski, M. A., Yoshimori, T., & Brumell, J. H. (2006). Autophagy controls Salmonella infection in response to damage to the Salmonella-containing vacuole. *The Journal of Biological Chemistry*, *281*(16), 11374–83. http://doi.org/10.1074/jbc.M509157200

Bjørkøy, G., Lamark, T., Brech, A., Outzen, H., Perander, M., Overvatn, A., … Johansen, T. (2005). p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *The Journal of Cell Biology*, *171*(4), 603–14. http://doi.org/10.1083/jcb.200507002

Boada-Romero, E., Letek, M., Fleischer, A., Pallauf, K., Ramón-Barros, C., & Pimentel-Muiños, F. X. (2013). TMEM59 defines a novel ATG16L1-binding motif that promotes local activation of LC3. *The EMBO Journal*, *32*(4), 566–82. http://doi.org/10.1038/emboj.2013.8

Bollin, F., Dechavanne, V., & Chevalet, L. (2011). Design of Experiment in CHO and HEK transient transfection condition optimization. *Protein Expression and Purification*, *78*(1), 61–8. http://doi.org/10.1016/j.pep.2011.02.008

Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., … Inohara, N. (2003). An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunology*, *4*(7), 702–7. http://doi.org/10.1038/ni945

Chaturvedi, A., Dorward, D., & Pierce, S. K. (2008). The B cell receptor governs the subcellular location of Toll-like receptor 9 leading to hyperresponses to DNA-containing antigens. *Immunity*, *28*(6), 799–809. http://doi.org/10.1016/j.immuni.2008.03.019

Chauhan, S., Mandell, M. A., & Deretic, V. (2015). IRGM Governs the Core Autophagy Machinery to Conduct Antimicrobial Defense. *Molecular Cell*, *58*(3), 507–521. http://doi.org/10.1016/j.molcel.2015.03.020

Chen, D., Fan, W., Lu, Y., Ding, X., Chen, S., & Zhong, Q. (2012). A mammalian autophagosome maturation mechanism mediated by TECPR1 and the Atg12-Atg5 conjugate. *Molecular Cell*, *45*(5), 629–41. http://doi.org/10.1016/j.molcel.2011.12.036

Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, *50*(Pt 5), 760–3. http://doi.org/10.1107/S0907444994003112

Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., … Simmons, A. (2010). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Medicine*, *16*(1), 90–97. http://doi.org/10.1038/nm.2069

Cuervo, A. M., & Wong, E. (2014). Chaperone-mediated autophagy: roles in disease and aging. *Cell Research*, *24*(1), 92–104. http://doi.org/10.1038/cr.2013.153

Cummings, M. D., Farnum, M. A., & Nelen, M. I. (2006). Universal screening methods and applications of ThermoFluor. *Journal of Biomolecular Screening*, *11*(7), 854–63. http://doi.org/10.1177/1087057106292746

Delgado, M. A., Elmaoued, R. A., Davis, A. S., Kyei, G., & Deretic, V. (2008). Toll-like receptors control autophagy. *The EMBO Journal*, *27*(7), 1110–21. http://doi.org/10.1038/emboj.2008.31

Ding, W.-X., Ni, H.-M., Li, M., Liao, Y., Chen, X., Stolz, D. B., … Yin, X.-M. (2010). Nix is critical to two distinct phases of mitophagy, reactive oxygen species-mediated autophagy induction and Parkin-ubiquitin-p62-mediated mitochondrial priming. *The Journal of Biological Chemistry*, *285*(36), 27879–90. http://doi.org/10.1074/jbc.M110.119537

Dooley, H. C., Razi, M., Polson, H. E. J., Girardin, S. E., Wilson, M. I., & Tooze, S. A. (2014). WIPI2 links LC3 conjugation with PI3P, autophagosome formation, and pathogen clearance by recruiting Atg12-5-16L1. *Molecular Cell*, *55*(2), 238–52. http://doi.org/10.1016/j.molcel.2014.05.021

Dorn, B. R., Dunn, W. A., & Progulske-Fox, A. (2001). Porphyromonas gingivalis Traffics to Autophagosomes in Human Coronary Artery Endothelial Cells. *Infection and Immunity*, *69*(9), 5698–5708. http://doi.org/10.1128/IAI.69.9.5698-5708.2001

Dunn, J. (1990). Studies on the mechanisms of autophagy: formation of the vacuole. *Journal of Cell Biology*, *110*(June), 1923–1933. http://doi.org/10.1083/jcb.110.6.1923

Dupont, N., Lacas-Gervais, S., Bertout, J., Paz, I., Freche, B., Van Nhieu, G. T., … Lafont, F. (2009). Shigella phagocytic vacuolar membrane remnants participate in the cellular response to pathogen invasion and are regulated by autophagy. *Cell Host & Microbe*, *6*(2), 137–49. http://doi.org/10.1016/j.chom.2009.07.005

Durocher, Y., Perret, S., & Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Research*, *30*(2), E9.

Eisenberg, D., Schwarz, E., Komaromy, M., & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology*, *179*(1), 125–142. http://doi.org/10.1016/0022-2836(84)90309-7

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2126–32. http://doi.org/10.1107/S0907444904019158

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501. http://doi.org/10.1107/S0907444910007493

Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, *357*(2), 289–98. http://doi.org/10.1016/j.ab.2006.07.027

Eskelinen, E. L. (2005). Maturation of autophagic vacuoles in Mammalian cells. *Autophagy*, *1*(1), 1–10. http://doi.org/10.4161/auto.1.1.1270

Feng, Y., He, D., Yao, Z., & Klionsky, D. J. (2014). The machinery of macroautophagy. *Cell Research*, *24*(1), 24–41. http://doi.org/10.1038/cr.2013.168

Fong, H. K., Hurley, J. B., Hopkins, R. S., Miake-Lye, R., Johnson, M. S., Doolittle, R. F., & Simon, M. I. (1986). Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(7), 2162–6.

Fujioka, Y., Noda, N. N., Nakatogawa, H., Ohsumi, Y., & Inagaki, F. (2010). Dimeric coiled-coil structure of Saccharomyces cerevisiae Atg16 and its functional significance in autophagy. *The Journal of Biological Chemistry*, *285*(2), 1508–15. http://doi.org/10.1074/jbc.M109.053520

Fujita, N., Itoh, T., Omori, H., Fukuda, M., Noda, T., & Yoshimori, T. (2008). The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. *Molecular Biology of the Cell*, *19*(5), 2092–100. http://doi.org/10.1091/mbc.E07-12-1257

Gammoh, N., Florey, O., Overholtzer, M., & Jiang, X. (2013). Interaction between FIP200 and ATG16L1 distinguishes ULK1 complex-dependent and -independent autophagy. *Nature Structural & Molecular Biology*, *20*(2), 144–9. http://doi.org/10.1038/nsmb.2475

Geisler, S., Holmström, K. M., Skujat, D., Fiesel, F. C., Rothfuss, O. C., Kahle, P. J., & Springer, W. (2010). PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biology*, *12*(2), 119–31. http://doi.org/10.1038/ncb2012

Geng, J., & Klionsky, D. J. (2008). The Atg8 and Atg12 ubiquitin-like conjugation systems in macroautophagy. "Protein modifications: beyond the usual suspects" review series. *EMBO Reports*, *9*(9), 859–64. http://doi.org/10.1038/embor.2008.163

Girardin, S. E., Boneca, I. G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., … Sansonetti, P. J. (2003). Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *The Journal of Biological Chemistry*, *278*(11), 8869–72. http://doi.org/10.1074/jbc.C200651200

Girardin, S. E., Travassos, L. H., Hervé, M., Blanot, D., Boneca, I. G., Philpott, D. J., … Mengin-Lecreulx, D. (2003). Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. *The Journal of Biological Chemistry*, *278*(43), 41702–8. http://doi.org/10.1074/jbc.M307198200

Graham, F. L., Smiley, J., Russell, W. C., & Nairn, R. (1977). Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *Journal of General Virology*, *36*(1), 59–72. http://doi.org/10.1099/0022-1317-36-1-59

Gutierrez, M. G., Master, S. S., Singh, S. B., Taylor, G. A., Colombo, M. I., & Deretic, V. (2004). Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell*, *119*(6), 753–66. http://doi.org/10.1016/j.cell.2004.11.038

Hampe, J., Cuthbert, A., Croucher, P. J., Mirza, M. M., Mascheretti, S., Fisher, S., … Mathew, C. G. (2001). Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet (London, England)*, *357*(9272), 1925–8. http://doi.org/10.1016/S0140-6736(00)05063-7

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., … Schreiber, S. (2006). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics*, *39*(2), 207–211. http://doi.org/10.1038/ng1954

He, Y. J., McCall, C. M., Hu, J., Zeng, Y., & Xiong, Y. (2006). D. *Genes & Development*, *20*(21), 2949–54. http://doi.org/10.1101/gad.1483206

Hemelaar, J., Lelyveld, V. S., Kessler, B. M., & Ploegh, H. L. (2003). A single protease, Apg4B, is specific for the autophagy-related ubiquitin-like proteins GATE-16, MAP1-LC3, GABARAP, and Apg8L. *The Journal of Biological Chemistry*, *278*(51), 51841–50. http://doi.org/10.1074/jbc.M308762200

Hosokawa, N., Hara, T., Kaizuka, T., Kishi, C., Takamura, A., Miura, Y., … Mizushima, N. (2009). Nutrient-dependent mTORC1 Association with the ULK1-Atg13-FIP200 Complex Required for Autophagy. *Molecular Biology of the Cell*, *20*(7), 1981–1991. http://doi.org/10.1091/mbc.E08-12-1248

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J. P., Belaiche, J., … Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, *411*(6837), 599–603. http://doi.org/10.1038/35079107

Ichimura, Y., Kirisako, T., Takao, T., Satomi, Y., Shimonishi, Y., Ishihara, N., … Ohsumi, Y. (2000). A ubiquitin-like system mediates protein lipidation. *Nature*, *408*(6811), 488–92. http://doi.org/10.1038/35044114

Inohara, N., Koseki, T., Lin, J., del Peso, L., Lucas, P. C., Chen, F. F., … Núñez, G. (2000). An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. *The Journal of Biological Chemistry*, *275*(36), 27823–31. http://doi.org/10.1074/jbc.M003415200

Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., … Nuñez, G. (2003). Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *The Journal of Biological Chemistry*, *278*(8), 5509–12. http://doi.org/10.1074/jbc.C200673200

Isakson, P., Holland, P., & Simonsen, A. (2013). The role of ALFY in selective autophagy. *Cell Death and Differentiation*, *20*(1), 12–20. http://doi.org/10.1038/cdd.2012.66

Ishibashi, K., Fujita, N., Kanno, E., Omori, H., Yoshimori, T., Itoh, T., & Fukuda, M. (2011). Atg16L2, a novel isoform of mammalian Atg16L that is not essential for canonical autophagy despite forming an Atg12–5-16L2 complex. *Autophagy*, *7*(12), 1500–13.

Itoh, T., Fujita, N., Kanno, E., Yamamoto, A., Yoshimori, T., & Fukuda, M. (2008). Golgi-resident Small GTPase Rab33B Interacts with Atg16L and Modulates Autophagosome Formation, *19*(July), 2916–2925. http://doi.org/10.1091/mbc.E07

Jäger, S., Bucci, C., Tanida, I., Ueno, T., Kominami, E., Saftig, P., & Eskelinen, E.-L. (2004). Role for Rab7 in maturation of late autophagic vacuoles. *Journal of Cell Science*, *117*(Pt 20), 4837–48. http://doi.org/10.1242/jcs.01370

Johansen, T., & Lamark, T. (2014). Selective autophagy mediated by autophagic adapter proteins. *Autophagy*, *7*(3), 279–296. http://doi.org/10.4161/auto.7.3.14487

Kabeya, Y., Mizushima, N., Ueno, T., Yamamoto, A., Kirisako, T., Noda, T., … Yoshimori, T. (2000). LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *The EMBO Journal*, *19*(21), 5720–8. http://doi.org/10.1093/emboj/19.21.5720

Kabeya, Y., Mizushima, N., Yamamoto, A., Oshitani-Okamoto, S., Ohsumi, Y., & Yoshimori, T. (2004). LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *Journal of Cell Science*, *117*(Pt 13), 2805–12. http://doi.org/10.1242/jcs.01131

Kabsch, W. (2010). XDS. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 125–32. http://doi.org/10.1107/S0907444909047337

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science : A Publication of the Protein Society*, *12*(9), 1865–1871. http://doi.org/10.1110/ps.0350503

Kawai, T., & Akira, S. (2006). Innate immune recognition of viral infection. *Nature Immunology*, *7*(2), 131–7. http://doi.org/10.1038/ni1303

Kim, P. K., Hailey, D. W., Mullen, R. T., & Lippincott-Schwartz, J. (2008). Ubiquitin signals autophagic degradation of cytosolic proteins and peroxisomes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(52), 20567–74. http://doi.org/10.1073/pnas.0810611105

Kirkin, V., Lamark, T., Sou, Y.-S., Bjørkøy, G., Nunn, J. L., Bruun, J.-A., … Johansen, T. (2009). A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. *Molecular Cell*, *33*(4), 505–16. http://doi.org/10.1016/j.molcel.2009.01.020

Kuballa, P., Huett, A., Rioux, J. D., Daly, M. J., & Xavier, R. J. (2008). Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. *PloS One*, *3*(10), e3391. http://doi.org/10.1371/journal.pone.0003391

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685. http://doi.org/10.1038/227680a0

Lassen, K. G., Kuballa, P., Conway, K. L., Patel, K. K., Becker, C. E., Peloquin, J. M., … Xavier, R. J. (2014). Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(21), 7741–6. http://doi.org/10.1073/pnas.1407001111

Lee, J., & Zhou, P. (2007). DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. *Molecular Cell*, *26*(6), 775–80. http://doi.org/10.1016/j.molcel.2007.06.001

Levine, B., Mizushima, N., & Virgin, H. W. (2011). Autophagy in immunity and inflammation. *Nature*, *469*(7330), 323–35. http://doi.org/10.1038/nature09782

Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., … Georges, M. (2007). Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genetics*, *3*(4), e58. http://doi.org/10.1371/journal.pgen.0030058

Lipinski, M. M., Hoffman, G., Ng, A., Zhou, W., Py, B. F., Hsu, E., … Yuan, J. (2010). A genome-wide siRNA screen reveals multiple mTORC1 independent signaling pathways regulating autophagy under normal nutritional conditions. *Developmental Cell*, *18*(6), 1041–52. http://doi.org/10.1016/j.devcel.2010.05.005

Liu, J. Z., & Anderson, C. A. (2014). Genetic studies of Crohn's disease: past, present and future. *Best Practice & Research. Clinical Gastroenterology*, *28*(3), 373–86. http://doi.org/10.1016/j.bpg.2014.04.009

Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, *33*(2), 491–7.

Mercer, C. a., Kaliappan, A., & Dennis, P. B. (2009). A novel, human Atg13 binding protein, Atg101, interacts with ULK1 and is essential for macroautophagy. *Autophagy*, *5*(5), 649–662. http://doi.org/10.4161/auto.5.5.8249

Mizushima, N. (2003). Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *Journal of Cell Science*, *116*(9), 1679–1688. http://doi.org/10.1242/jcs.00381

Mizushima, N., Noda, T., & Ohsumi, Y. (1999). Apg16p is required for the function of the Apg12p-Apg5p conjugate in the yeast autophagy pathway. *The EMBO Journal*, *18*(14), 3888–96. http://doi.org/10.1093/emboj/18.14.3888

Mizushima, N., Noda, T., Yoshimori, T., Tanaka, Y., Ishii, T., George, M. D., … Ohsumi, Y. (1998). A protein conjugation system essential for autophagy. *Nature*, *395*(6700), 395–8. http://doi.org/10.1038/26506

Mizushima, N., Yamamoto, A., Hatano, M., Kobayashi, Y., Kabeya, Y., Suzuki, K., … Yoshimori, T. (2001). Dissection of Autophagosome Formation Using Apg5-Deficient Mouse Embryonic Stem Cells. *The Journal of Cell Biology*, *152*(4), 657–668. http://doi.org/10.1083/jcb.152.4.657

Mizushima, N., & Yoshimori, T. (2014). How to Interpret LC3 Immunoblotting. *Autophagy*, *3*(6), 542–545. http://doi.org/10.4161/auto.4600

Moreau, K., Ravikumar, B., Renna, M., Puri, C., & Rubinsztein, D. C. (2011). Autophagosome precursor maturation requires homotypic fusion. *Cell*, *146*(2), 303–17. http://doi.org/10.1016/j.cell.2011.06.023

Murthy, A., Li, Y., Peng, I., Reichelt, M., Katakam, A. K., Noubade, R., … van Lookeren Campagne, M. (2014). A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature*, *506*(7489), 456–62. http://doi.org/10.1038/nature13044

Nair, U., Jotwani, A., Geng, J., Gammoh, N., Richerson, D., Yen, W.-L., … Klionsky, D. J. (2011). SNARE proteins are required for macroautophagy. *Cell*, *146*(2), 290–302. http://doi.org/10.1016/j.cell.2011.06.022

Nakagawa, I., Amano, A., Mizushima, N., Yamamoto, A., Yamaguchi, H., Kamimoto, T., … Yoshimori, T. (2004). Autophagy defends cells against invading group A Streptococcus. *Science (New York, N.Y.)*, *306*(5698), 1037–40. http://doi.org/10.1126/science.1103966

Nakatogawa, H., Suzuki, K., Kamada, Y., & Ohsumi, Y. (2009). Dynamics and diversity in autophagy mechanisms: lessons from yeast. *Nature Reviews. Molecular Cell Biology*, *10*(7), 458–67. http://doi.org/10.1038/nrm2708

Naser, S. a, Arce, M., Khaja, A., Fernandez, M., Naser, N., Elwasila, S., & Thanigachalam, S. (2012). Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World Journal of Gastroenterology : WJG*, *18*(5), 412–24. http://doi.org/10.3748/wjg.v18.i5.412

Neer, E. J., Schmidt, C. J., Nambudripad, R., & Smith, T. F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature*, *371*(6495), 297–300. http://doi.org/10.1038/371297a0

Nishimura, T., Kaizuka, T., Cadwell, K., Sahani, M. H., Saitoh, T., Akira, S., … Mizushima, N. (2013). FIP200 regulates targeting of Atg16L1 to the isolation membrane. *EMBO Reports*, *14*(3), 284–91. http://doi.org/10.1038/embor.2013.6

Ogawa, M., Yoshimori, T., Suzuki, T., Sagara, H., Mizushima, N., & Sasakawa, C. (2005). Escape of intracellular Shigella from autophagy. *Science (New York, N.Y.)*, *307*(5710), 727–31. http://doi.org/10.1126/science.1106036

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., … Cho, J. H. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, *411*(6837), 603–6. http://doi.org/10.1038/35079114

Okamoto, K. (2014). Organellophagy: Eliminating cellular building blocks via selective autophagy. *Journal of Cell Biology*, *205*(4), 435–445. http://doi.org/10.1083/jcb.201402054

Orsi, A., Razi, M., Dooley, H. C., Robinson, D., Weston, A. E., Collinson, L. M., & Tooze, S. A. (2012). Dynamic and transient interactions of Atg9 with autophagosomes, but not membrane integration, are required for autophagy. *Molecular Biology of the Cell*, *23*(10), 1860–73. http://doi.org/10.1091/mbc.E11-09-0746

Orvedahl, A., MacPherson, S., Sumpter, R., Tallóczy, Z., Zou, Z., & Levine, B. (2010). Autophagy protects against Sindbis virus infection of the central nervous system. *Cell Host & Microbe*, *7*(2), 115–27. http://doi.org/10.1016/j.chom.2010.01.007

Otomo, C., Metlagel, Z., Takaesu, G., & Otomo, T. (2013). Structure of the human ATG12~ATG5 conjugate required for LC3 lipidation in autophagy. *Nature Structural & Molecular Biology*, *20*(1), 59–66. http://doi.org/10.1038/nsmb.2431

Parkes, M., Barrett, J. C., Prescott, N. J., Tremelling, M., Anderson, C. A., Fisher, S. A., … Mathew, C. G. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn’s disease susceptibility. *Nature Genetics*, *39*(7), 830–2. http://doi.org/10.1038/ng2061

Parkhouse, R., Ebong, I.-O., Robinson, C. V, & Monie, T. P. (2013). The N-terminal region of the human autophagy protein ATG16L1 contains a domain that folds into a helical structure consistent with formation of a coiled-coil. *PloS One*, *8*(9), e76237. http://doi.org/10.1371/journal.pone.0076237

Perrin, A. J., Jiang, X., Birmingham, C. L., So, N. S. Y., & Brumell, J. H. (2004). Recognition of bacteria in the cytosol of Mammalian cells by the ubiquitin system. *Current Biology : CB*, *14*(9), 806–11. http://doi.org/10.1016/j.cub.2004.04.033

Pizarro-Cerda, J., Meresse, S., Parton, R. G., van der Goot, G., Sola-Landa, A., Lopez-Goni, I., … Gorvel, J.-P. (1998). Brucella abortus Transits through the Autophagic Pathway and Replicates in the Endoplasmic Reticulum of Nonprofessional Phagocytes. *Infect. Immun.*, *66*(12), 5711–5724.

Polson, H. E. J., de Lartigue, J., Rigden, D. J., Reedijk, M., Urbé, S., Clague, M. J., & Tooze, S. A. (2010). Mammalian Atg18 (WIPI2) localizes to omegasome-anchored phagophores and positively regulates LC3 lipidation. *Autophagy*, *6*(4), 506–22. http://doi.org/10.4161/auto.6.4.11863

Ramjeet, M., Hussey, S., Philpott, D. J., & Travassos, L. H. (2010). “Nodophagy”: New crossroads in Crohn disease pathogenesis. *Gut Microbes*, *1*(5), 307–315. http://doi.org/10.4161/gmic.1.5.13295

Rupp, B. (2010). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. Retrieved August 13, 2015, from http://www.amazon.com/Biomolecular-Crystallography-Principles-Application-Structural/dp/0815340818

Russell, R. B., Sasieni, P. D., & Sternberg, M. J. (1998). Supersites within superfolds. Binding site similarity in the absence of homology. *Journal of Molecular Biology*, *282*(4), 903–18. http://doi.org/10.1006/jmbi.1998.2043

Russell, R. C., Tian, Y., Yuan, H., Park, H. W., Chang, Y.-Y., Kim, J., … Guan, K.-L. (2013). ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. *Nature Cell Biology*, *15*(7), 741–50. http://doi.org/10.1038/ncb2757

Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J., & Verdine, G. L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nature Structural & Molecular Biology*, *13*(8), 704–12. http://doi.org/10.1038/nsmb1119

Sanjuan, M. A., Dillon, C. P., Tait, S. W. G., Moshiach, S., Dorsey, F., Connell, S., … Green, D. R. (2007). Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature*, *450*(7173), 1253–7. http://doi.org/10.1038/nature06421

Sartor, R. B. (2005). Role of commensal enteric bacteria in the pathogenesis of immune-mediated intestinal inflammation: lessons from animal models and implications for translational research. *Journal of Pediatric Gastroenterology and Nutrition*, *40 Suppl 1*(April), S30–S31. http://doi.org/10.1097/00005176-200504001-00018

Schmid, D., & Münz, C. (2007). Innate and adaptive immunity through autophagy. *Immunity*, *27*(1), 11–21. http://doi.org/10.1016/j.immuni.2007.07.004

Scrima, A., Fischer, E. S., Lingaraju, G. M., Böhm, K., Cavadini, S., & Thomä, N. H. (2011). Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! *FEBS Letters*, *585*(18), 2818–25. http://doi.org/10.1016/j.febslet.2011.04.064

Scrima, A., Konícková, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., … Thomä, N. H. (2008). Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, *135*(7), 1213–23. http://doi.org/10.1016/j.cell.2008.10.045

Sheldrick, G. M. (2002). Macromolecular phasing with SHELXE. *Zeitschrift Für Kristallographie*, *217*(12-2002), 644–650. http://doi.org/10.1524/zkri.217.12.644.20662

Sheldrick, G. M. (2007). A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, *64*(1), 112–122. http://doi.org/10.1107/S0108767307043930

Shin, D.-M., Yuk, J.-M., Lee, H.-M., Lee, S.-H., Son, J. W., Harding, C. V, … Jo, E.-K. (2010). Mycobacterial lipoprotein activates autophagy via TLR2/1/CD14 and a functional vitamin D receptor signalling. *Cellular Microbiology*, *12*(11), 1648–65. http://doi.org/10.1111/j.1462-5822.2010.01497.x

Singh, S. B., Davis, A. S., Taylor, G. A., & Deretic, V. (2006). Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science (New York, N.Y.)*, *313*(5792), 1438–41. http://doi.org/10.1126/science.1129577

Singh, S. B., Ornatowski, W., Vergne, I., Naylor, J., Delgado, M., Roberts, E., … Deretic, V. (2010). Human IRGM regulates autophagy and cell-autonomous immunity functions through mitochondria. *Nature Cell Biology*, *12*(12), 1154–65. http://doi.org/10.1038/ncb2119

Smith, T. F., Gaitatzes, C., Saxena, K., & Neer, E. J. (1999). The WD repeat: a common architecture for diverse functions. *Trends in Biochemical Sciences*, *24*(5), 181–185. http://doi.org/10.1016/S0968-0004(99)01384-5

Sorbara, M. T., Ellison, L. K., Ramjeet, M., Travassos, L. H., Jones, N. L., Girardin, S. E., & Philpott, D. J. (2013). The protein ATG16L1 suppresses inflammatory cytokines induced by the intracellular sensors Nod1 and Nod2 in an autophagy-independent manner. *Immunity*, *39*(5), 858–73. http://doi.org/10.1016/j.immuni.2013.10.013

Srimathi, T., Robbins, S. L., Dubas, R. L., Hasegawa, M., Inohara, N., & Park, Y. C. (2008). Monomer/dimer transition of the caspase-recruitment domain of human Nod1. *Biochemistry*, *47*(5), 1319–25. http://doi.org/10.1021/bi7016602

Srimathi, T., Robbins, S. L., Dubas, R. L., Seo, J. H., & Park, Y. C. (2007). Purification, crystallization and preliminary crystallographic characterization of the caspase-recruitment domain of human Nod1. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *63*(Pt 1), 21–3. http://doi.org/10.1107/S1744309106051955

Stein, N. (2008). CHAINSAW : a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, *41*(3), 641–643. http://doi.org/10.1107/S0021889808006985

Stirnimann, C. U., Petsalaki, E., Russell, R. B., & Müller, C. W. (2010). WD40 proteins propel cellular networks. *Trends in Biochemical Sciences*, *35*(10), 565–74. http://doi.org/10.1016/j.tibs.2010.04.003

Sun, Q., Zhang, J., Fan, W., Wong, K. N., Ding, X., Chen, S., & Zhong, Q. (2011). The RUN domain of rubicon is important for hVps34 binding, lipid kinase inhibition, and autophagy suppression. *The Journal of Biological Chemistry*, *286*(1), 185–91. http://doi.org/10.1074/jbc.M110.126425

Suzuki, K., Kirisako, T., Kamada, Y., Mizushima, N., Noda, T., & Ohsumi, Y. (2001). The pre-autophagosomal structure organized by concerted functions of APG genes is essential for autophagosome formation. *The EMBO Journal*, *20*(21), 5971–81. http://doi.org/10.1093/emboj/20.21.5971

Suzuki, N. N., Yoshimoto, K., Fujioka, Y., Ohsumi, Y., & Inagaki, F. (2005). The crystal structure of plant ATG12 and its biological implication in autophagy. *Autophagy*, *1*(2), 119–126. http://doi.org/10.4161/auto.1.2.1859

Sventoraityte, J., Zvirbliene, A., Franke, A., Kwiatkowski, R., Kiudelis, G., Kupcinskas, L., & Schreiber, S. (2010). NOD2, IL23R and ATG16L1 polymorphisms in Lithuanian patients with inflammatory bowel disease. *World Journal of Gastroenterology : WJG*, *16*(3), 359–64.

Tallóczy, Z., Herbert Virgin, I., & Levine, B. (2005). PKR-Dependent Xenophagic Degradation of Herpes Simplex Virus Type 1. *Autophagy*.

Tanida, I., Sou, Y., Minematsu-Ikeguchi, N., Ueno, T., & Kominami, E. (2006). Atg8L/Apg8L is the fourth mammalian modifier of mammalian Atg8 conjugation mediated by human Atg4B, Atg7 and Atg3. *The FEBS Journal*, *273*(11), 2553–62. http://doi.org/10.1111/j.1742-4658.2006.05260.x

Thurston, T. L. M., Ryzhakov, G., Bloor, S., von Muhlinen, N., & Randow, F. (2009). The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. *Nature Immunology*, *10*(11), 1215–21. http://doi.org/10.1038/ni.1800

Thurston, T. L. M., Wandel, M. P., von Muhlinen, N., Foeglein, A., & Randow, F. (2012). Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature*, *482*(7385), 414–8. http://doi.org/10.1038/nature10744

Travassos, L. H., Carneiro, L. a M., Ramjeet, M., Hussey, S., Kim, Y.-G., Magalhães, J. G., … Philpott, D. J. (2010). Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunology*, *11*(1), 55–62. http://doi.org/10.1038/ni.1823

Travassos, L. H., Carneiro, L. A. M., Girardin, S., & Philpott, D. J. (2010). Nod proteins link bacterial sensing and autophagy. *Autophagy*, 409–411. http://doi.org/10.1038/ni.1823.utophagy

Urzhumtseva, L., Afonine, P. V, Adams, P. D., & Urzhumtsev, A. (2009). Crystallographic model quality at a glance. *Acta Crystallographica. Section D, Biological Crystallography*, *65*(Pt 3), 297–300. http://doi.org/10.1107/S0907444908044296

Vaughn, J. L., Goodwin, R. H., Tompkins, G. J., & McCawley, P. (1977). The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro*, *13*(4), 213–7.

Vermeire, S., Wild, G., Kocher, K., Cousineau, J., Dufresne, L., Bitton, A., … Rioux, J. D. (2002). CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. *American Journal of Human Genetics*, *71*(1), 74–83. http://doi.org/10.1086/341124

Wang, H., Sun, H.-Q., Zhu, X., Zhang, L., Albanesi, J., Levine, B., & Yin, H. (2015). GABARAPs regulate PI4P-dependent autophagosome:lysosome fusion. *Proceedings of the National Academy of Sciences*, *112*(22), 201507263. http://doi.org/10.1073/pnas.1507263112

Wasilko, D. J., Lee, S. E., Stutzman-Engwall, K. J., Reitz, B. A., Emmons, T. L., Mathis, K. J., … Fischer, H. D. (2009). The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. *Protein Expression and Purification*, *65*(2), 122–32. http://doi.org/10.1016/j.pep.2009.01.002

Weichenberger, C. X., & Rupp, B. (2014). Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallographica. Section D, Biological Crystallography*, *70*(Pt 6), 1579–88. http://doi.org/10.1107/S1399004714005550

Wickham, T. J., & Nemerow, G. R. Optimization of growth methods and recombinant protein production in BTI-Tn-5B1-4 insect cells using the baculovirus expression system. *Biotechnology Progress*, *9*(1), 25–30. http://doi.org/10.1021/bp00019a004

Wild, P., Farhan, H., McEwan, D. G., Wagner, S., Rogov, V. V, Brady, N. R., … Dikic, I. (2011). Phosphorylation of the autophagy receptor optineurin restricts Salmonella growth. *Science (New York, N.Y.)*, *333*(6039), 228–33. http://doi.org/10.1126/science.1205405

Xu, C., & Min, J. (2011). Structure and function of WD40 domain proteins. *Protein and Cell*, *2*(3), 202–214. http://doi.org/10.1007/s13238-011-1018-1

Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J. L., & Tong, L. (2000). Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature*, *408*(6808), 111–5. http://doi.org/10.1038/35040600

Yang, Z., & Klionsky, D. J. (2010). Mammalian autophagy: core molecular machinery and signaling regulation. *Current Opinion in Cell Biology*, *22*(2), 124–31. http://doi.org/10.1016/j.ceb.2009.11.014

Yoshikawa, Y., Ogawa, M., Hain, T., Yoshida, M., Fukumatsu, M., Kim, M., … Sasakawa, C. (2009). Listeria monocytogenes ActA-mediated escape from autophagic recognition. *Nature Cell Biology*, *11*(10), 1233–40. http://doi.org/10.1038/ncb1967

Young, A. R. J., Chan, E. Y. W., Hu, X. W., Köchl, R., Crawshaw, S. G., High, S., … Tooze, S. A. (2006). Starvation and ULK1-dependent cycling of mammalian Atg9 between the TGN and endosomes. *Journal of Cell Science*, *119*(Pt 18), 3888–900. http://doi.org/10.1242/jcs.03172

Zheng, Y. T., Shahnazari, S., Brech, A., Lamark, T., Johansen, T., & Brumell, J. H. (2009). The adaptor protein p62/SQSTM1 targets invading bacteria to the autophagy pathway. *Journal of Immunology (Baltimore, Md. : 1950)*, *183*(9), 5909–16. http://doi.org/10.4049/jimmunol.0900441

Acknowledgements

It is my pleasure to acknowledge the roles of several people who were essential for completion of my Ph.D. work.

First and foremost, I want to thank my supervisor Dr. Andrea Scrima for his support and guidance during my thesis. I truly enjoyed working in the environment he created in the group, which allows personal growth, original thinking and initiative. Dr. Scrima's professional advice and ideas and above all his stoic patience and understanding are enormously appreciated.

I would like to acknowledge helpful suggestions and assistance from my committee members: Prof. Dr. Joop van den Heuvel and Prof. Dr. Luka Cicin-Sain.

Additionally, I would like to thank Prof. Dr. Michael Steinert and Prof. Dr. Wulf Blankenfeldt for agreeing to be on my examination committee, and Prof. Dr. Michael Hust for serving as the head of my Ph.D. defence commission.

I specially want to thank Dr. Joop van den Heuvel and all the members of the RPEX group for providing the platform for insect and mammalian cell culture, their help and expert advice.

These acknowledgements would not be complete without a special mention of all the members of the SBAU group: Petra Völler, Stefan Leupold, Archna, Dr. Stefan Schmelz, Pooja Sadana and Dr. Caroline Behrens. It was a great pleasure working with this group and I appreciate all their help and support, in and out of the lab, and their good humor and friendship.

I have to express my heartfelt gratitude to all the members of MOSB and SFPR for their friendship and professional advice. Special thanks to Dr. Lukat for the help with figures and PyMOL and to Dr. Krausse for inexhaustible crystallography knowledge and relentless enthusiasm for discussion.

My deepest appreciation goes to my parents for their undying faith in me. And to Robert, for all his love.

Curriculum Vitae

PERSÖNLICHE ANGABEN

Geburtsdatum: 11.12.1983

Geburtsort: Podgorica, Montenegro

AUSBILDUNG

2001 – 2006	University of Wisconsin – Parkside	Kenosha, WI, USA
	B.S., Biological Sciences	
2007 – 2008	University of Wisconsin – Parkside	Kenosha, WI, USA
	Erwerb aller Credit Points für den Abschluss M.Sc., Applied Molecular Biology	
2009 – 2011	Eidgenössiche Technische Hochschule	Zürich, Schweiz
	M.Sc., Mikrobiologie und Immunologie	
	Die Masterarbeit (Dauer 6 Monate) wurde in der Forschungsgruppe von Dr. Christian Münz absolviert und behandelte den Nachweis der Relevanz von Makroautophagie und/oder durch Makroautophagie vermittelte Phagocytose zur Kontrolle von Infektionen durch *Candida albicans*. Die Arbeit resultierte in einer Veröffentlichung.	
2012-2015	Ph.D. Kandidat TU Braunschweig	Braunschweig, Deutschland
	Die derzeitige Arbeit behandelt die Aufklärung der Struktur von menschlichem ATG16L1 unter Verwendung von Röntgenstrahlung. Tätigkeitsumfang: Konstrukt-Design, Klonierung, Expression von Proteinen von *E. coli* und Insektenzellen, Reinigung von Proteinen und Kristallisierung, Datensammlung und –prozessierung, Strukturaufklärung und –interpretation. Außerdem, Etablierung und Pflege von Kollaborationen mit drei Forschungsgruppen auf dem Campus.	

BERUFSERFAHRUNG

2004 – 2006	University of Wisconsin – Parkside	Kenosha, WI, USA
	Assistenz chemische Forschung	
2005 – 2007	BASF – The Chemical Company	Sturtevant, WI, USA
	Technische Assistentin	
2009 – 2010	Ostfalia HAW	Wolfenbüttel, Deutschland
	Assistenz Sprachunterricht	
Seit 2012	Helmholtz -Zentrum für Infektionsforschung	Braunschweig, Deutschland
	Erlangung des Grades Ph.D. in der Gruppe Structural Biology of Autophagy (SBAU)	

WEITERBILDUNG & BESONDERHEITEN

Besuchte Konferenzen

- Murnau Konferenz, 2012
- Heart of Europe of Crystallography, 2012 (HEC-15), 2013 (HEC-16), 2014 (HEC-17)
- NoRDI IV Symposium, 2013 und NoRDI V Symposium, 2014

Seminare

- International Presentation and Oration Skills, 2013
- International School of Biological Crystallography Granada, 2013
- International Teaming and Negotiation Skills, 2014
- Wissenschaftliches Schreiben, 2014
- FEI Cryo-TEM Workshop, 2015

Auszeichnungen

- Stipendium vom Dekan und Vizekanzler (URAP) für das Studienjahr 2005 und Frühling 2006
- Platz für bestes Poster, HZI Grad School Annual Retreat, 2011
- Platz für beste Präsentation, HZI Grad School PhD Symposium, 2014

Sprachkenntnisse

- Montenegrinisch: Muttersprache
- Englisch: Verhandlungssicher
- Deutsch: Gute Kenntnisse (Level B2)

VERÖFFENTLICHUNGEN

Romao S., Gasser N., Becker A.C., Guhl B., Bajagic M., Vanoaica D., Ziegler U., Roesler J., Dengjel J., Reichenbach J., Münz C. "Autophagy proteins stabilize pathogen-containing phagosomes for prolonged MHC II atigen processing.", J. Cell. Biol., 2013 Dec 9; 203(5):757-66

Kolb V. und Bajagic M. "The Maillard Reaction of the Meteoritic Amino Acids", Astrobiology, 6, 248 (2006).